

DOKTORI ÉRTEKEZÉS

**Extrém érték statisztikák
mintaméret problémái és
a galaxisok fényesség-eloszlása**

OZOGÁNY KATALIN

Témavezetők: Dr. Rácz Zoltán, az MTA rendes tagja
Dr. Györgyi Géza, PhD

Eötvös Loránd Tudományegyetem, Természettudományi Kar
Fizika Doktori Iskola

Iskolavezető: Dr. Csikor Ferenc, az MTA doktora

Statisztikus fizika, biológiai fizika és
kvantumrendszerek fizikája program

Programvezető: Dr. Kürti Jenő, az MTA doktora



Elméleti Fizika Tanszék
Eötvös Loránd Tudományegyetem
Budapest, 2012

Tartalomjegyzék

1. Bevezetés	5
2. Extrém érték statisztika renormálási csoport megközelítése	11
2.1. Bevezetés az iid változók extrém érték statisztikájába	12
2.2. A renormálási csoport transzformáció fixpontjai és az extrém határeloszlások	14
2.3. Perturbációk a fixpont körül és a véges méret skálázás	16
2.4. A véges méret korrekció és az őseloszlás kapcsolata	18
2.5. Az eloszlások és a véges méret korrekció alakja a két sztenderdizációban	19
2.6. Véges méret korrekció az FTG osztályban	21
3. Sorrend statisztika	31
3.1. $1/f^\alpha$ zajok sorrend statisztikája	32
3.1.1. Független változók, $\alpha = 0$	34
3.1.2. Korrelált változók, $\alpha \neq 0$	36
3.2. Az első és második legnagyobb érték várható távolsága iid változókra	39
3.2.1. A távolság határértéke általános esetben	40
3.2.2. A távolság véges méret korrekciója az FTG osztályra	42
4. Extrém érték statisztika véletlen mintaméret mellett	47
4.1. Az univerzális osztályok megjelenése két speciális esetben	48
4.1.1. Egyenletes mintaméret eloszlás	50
4.1.2. Exponenciális mintaméret eloszlás	51
4.2. Extrém határeloszlás és véges méret korrekció általános esetben	53
4.2.1. Skálázó mintaméret eloszlás	53

4.2.2. Nem skálázó mintaméret eloszlás	56
4.2.3. A változó mintaméret mint korrekció	57
4.3. Az első és második legnagyobb érték távolsága skálázó mintaméret eloszlásra	58
5. Galaxisok fényességének extrém statisztikája az SDSS adatai alapján	61
5.1. A használt adatbázis	62
5.2. Luminozítás eloszlás	66
5.3. Galaxisok számának eloszlása	72
5.4. A maximális luminozítások elméleti határeloszlása és véges méret korrekciója	76
5.5. Maximális luminozítás eloszlás meghatározása szimulációval	86
5.5.1. A szimuláció algoritmus	86
5.5.2. Maximális luminozítás eloszlás szimulálása	87
5.5.3. Nem skálázó mintaméret esetének vizsgálata szimulációkkal	89
5.6. Hisztogramok hibájának becslése	90
5.7. Gap statisztika és a Tremaine-Richstone arányszámok	93
6. Eredmények összefoglalása	99
A. Extrém határeloszlás két speciális esetben	103
A.1. Egyenletes mintaméret eloszlás, exponenciális őseloszlás	103
A.2. Exponenciális mintaméret eloszlás, exponenciális őseloszlás	106
B. Rövidítések	109
Irodalomjegyzék	111
Összefoglaló	117
Summary	119

1. fejezet

BEVEZETÉS

Az extrém értékek elmélete régóta ismert és kutatott terület a statisztikával kapcsolatos tudományágakban. Az utóbbi években nagyobb figyelem övezi, mivel az általa jellemzett ritkán előforduló, de extrém események, mint a jelentősebb földrengések, hurrikánok, árvizek vagy szárazságok, általában nagy horderejűek, és gyakran katasztrofális következménnyel járnak. Az extrém események így hatással vannak mindennapi életünkre is, ezért bír olyan nagy jelentőséggel statisztikájuk leírása.

Az extrém érték statisztika (extreme value statistics - EVS) az események eloszlása ismeretében próbál következtetni az extrém események eloszlására, és ezáltal lehetővé teszi előrejelzésüket. Eszközt ad olyan kérdések megválaszolására, mint hogy milyen magas gát véd meg adott valószínűséggel a 100 év alatt megjelenő legnagyobb árvíztől, ha rendelkezésünkre állnak a folyó vízállás adatai több száz évre visszamenőleg. Ehhez hasonló kérdések felmerülhetnek számos környezeti jelenséget illetően és az alkalmazások köre a kiterjedt adatbázisok megjelenésével egyre bővül.

Az eleinte csak matematikai érdeklődésre [1]-[4] számot tartó kérdéskör fokozatosan került be a mérnöki [5], környezettudományi, majd később a fizikai köztudatba. Jelentősége nem kétséges azokon az újabban fizikusok által is művelt területeken, mint a pénzügy és biztosítás [6], vagy olyan természeti csapások előrejelzésében, amelyek környezeti változók extrém értékével kapcsolhatók össze. Vizsgálhatóak a hőmérsékleti csúcsértékek, a csapadékmennyiség és szélerősség extrémális értékei, a tengerszint magassága vagy a földrengések statisztikája, így számos alkalmazásra került sor geológiai [7], hidrológiai [8], klimatológiai [9] kutatásokban, a klímaváltozás vizsgálatának és detektálásának terjedő eszköze [10, 11]. A fizika különböző területein is egyre több alkalmazás lát napvilágot, csak pár példát említve, rendezetlen rendszerek alacsony hőmérsékleti leírása [12], felületi fluktuációk [13]-[20],

random fragmentáció [21], kvantumgázok energiasűrűsége [22] kapcsán.

Az extrémumokkal kapcsolatban feltehető kérdések az alkalmazásokhoz hasonlóan sokrétűek. Az elmélet és a gyakorlat terén is a leggyakrabban felmerülő probléma egy változó extrém értékének valószínűségi eloszlása. A véletlen változó, ami általában egy fizikai mennyiséget takar, időbeli (de lehet tér- vagy sokaságbeli) megvalósult értékeit vizsgáljuk. A változó mennyiség előfordulási valószínűsége vagy elméleti úton meghatározható a háttérben működő fizikai folyamatok ismeretében, vagy egy mérési adatsorból becsülhető. A kérdés pedig a változó adott hosszúságú intervallumon belül előforduló maximális vagy minimális értékének valószínűségi eloszlása, statisztikai jellemzése. Mivel a minimum érték problémája, a változó előjelcseréje után, a maximummal ekvivalens probléma, ezért mi csak a maximumokról fogunk beszélni, de értelemszerűen az állítások a minimumra is vonatkoznak.

A maximális érték mellett érdekes lehet a k -adik legnagyobb érték eloszlása, vagy a nagyság szerint rendezett minta elemeinek egymáshoz viszonyított helyzete, mint például az első és második érték várható távolsága. Az alkalmazások terén inkább gyakorlati szempontból jelentős kérdések szoktak felmerülni, mint például milyen legyen a tervezési érték egy T visszatérési idővel jellemezhető jelenség kivédése érdekében, vagy egy adott küszöbérték esetén milyen gyakorisággal fordulnak elő szintátmetszések, vagy hogy milyen időközönként jelennek meg új rekordok egy idősorban.

Ami az extrém értékek eloszlását illeti, azok aszimptotikus viselkedése jól tisztázott a matematikai irodalomban [1, 2] arra a speciális esetre, amikor a vizsgált változók függetlenek és azonos eloszlásúak (independent identical distributed - iid). Ebben az esetben az eredeti változó eloszlása – az őseloszlás – analitikus ismerete esetén az N elemű minta maximumának eloszlása egyszerűen felírható. Az elmélet szerint az $N \rightarrow \infty$ limeszben univerzális osztályok jelennek meg, az őseloszlás, pontosabban annak nagy argumentumú aszimptotikus viselkedése határozza meg, hogy az extrém eloszlás melyik határeloszlás családjához fog tartani a lehetséges három közül (Fisher-Tippett-Gumbel, Fisher-Tippett-Fréchet, vagy Fisher-Tippett-Weibull).

Az elmélet által tisztázott ideális esettől az empirikusan megfigyelt eloszlások több okból is eltérhetnek. A gyakorlatban előforduló idősorok esetén a minta mérete általában nem túl nagy, így az empirikus, véges méretű, extrém érték eloszlása (amit a mérések alapján felvett hisztogrammal becsülünk) a határeloszlástól jelentősen különbözhet, és értelmezése csak a véges méret korrekció figyelembe vételével lehetséges. Ezenkívül a vizsgált változók gyakran korreláltak, és az iid elmélet érvényét veszti. Tovább bonyolítja a helyzetet, ha a minta mérete, amelyből a maximumot megfigyeljük, szintén egy véletlen változó. Mivel az EVS elméletének feltételezései a gyakorlat

során a legkritikább esetben teljesülnek, ezért elengedhetetlen részletesebben foglalkozni a fenti három effektussal, a véges mintaméret, a korreláció és a véletlen mintaméret hatásával. Ennek a három problémakörnek a vizsgálata adja a disszertáció fő témaköreit.

A véges méretből adódó korrekció (finite size correction - FSC) szerepe főként azért lényeges, mert az elméleti határeloszlásokhoz való konvergencia általában lassú. Ismert példája ennek a Fisher-Tippett-Gumbel (FTG) határeloszlás, amihez a konvergencia a leggyakrabban logaritmikus, $1/\ln N$ -nel arányos. Bár a konvergencia sebessége a matematikai irodalomnak köszönhetően elméleti alapokon eddig is ismert volt [23, 24, 25], a véges méret korrekció részletes, általunk kiterjesztett leírására csak nemrégiben került sor [26, 27]. Itt egy renormálási csoport bevezetésével meghatároztuk az FSC általános függvényalakját és az N mintamérettől függő amplitúdóját vezető rendben iid változókra. Ezen renormálási csoport transzformáció fixpont egyenletéből megkapható az egy paraméterrel (γ) jellemzett általános extrém határeloszlás. Más szóval az extrém határeloszlások a transzformáció fixpontjai, pontosabban egy γ -val parametrizált fix vonallal reprezentálható a függvényterben az általános extrém határeloszlás, amely magában foglalja a három határeloszlás családot. A véges méret korrekció alakja egy fixpont körüli lineárizációval kapható meg, általános formában. Univerzális vonás, hogy egy γ -val paraméterezett határeloszlás családon belül csupán egy paraméterrel (γ') jellemezhető az FSC elsőrendű tagja, amely γ' paramétert az őseloszlás aszimptotikus viselkedése határozza meg. Ennek az általános formulának a felhasználásával az őseloszlástól függően felírható a véges méret korrekció minden konkrét esetben. A dolgozat első fő témakörét az FSC leírása adja, és a renormálási csoport megközelítés eredményeinek ismertetése.

A véges méret korrekcióval kiegészített aszimptotikus iid elmélet az alkalmazások során felmerülő mintaméretek mellett is egy nagyon jó leírást ad független változókra. A korrelált változók esetéről azonban egyelőre keveset tudunk. Az EVS leggyakoribb alkalmazási területein ezzel szemben korrelált változókkal találkozunk. A legtöbb környezeti jelenséget – mint például a földrengés erősségét vagy a meteorológia és hidrológia terén vizsgált mennyiségeket – gyakran nagy fluktuációk jellemzik, ami erős térbeli ill. időbeli korreláltságra utal. Ebből következik, hogy nagy szükség van az EVS leírására korrelált változók esetén is. Néhány speciális esetben már születtek eredmények, amik azt mutatják, hogy különbséget kell tenni gyenge és erős korreláció között. Gyenge, hatványszerűen lecsengő, de véges integrállal rendelkező korreláció esetén a határeloszlás az iid elmélettel megegyezik [28], viszont numerikus eredmények azt mutatják, hogy a véges méret korrekció megváltozik [29]. Erős korreláció jelenlétében, egy analitikus úton kapott

eredmény szerint, a határeloszlás az iid esettől eltérő: [30, 31]-ben az $(1+1)$ dimenziós periodikus Edwards-Wilkinson felület magasságát nézték – ami egy erősen korrelált változó –, és ennek maximum értékének eloszlását vizsgálva az Airy-eloszlást kapták.

A korreláció hatásának kiterjedt vizsgálatára különösen alkalmasnak bizonyultak az $1/f^\alpha$ teljesítményspektrumú, gaussi fluktuációkat mutató, periodikus jelek. Az α paraméterrel a korreláció mértéke növelhető, $\alpha = 0$ a korrelálatlan, iid esetnek felel meg, $\alpha < 1$ mellett hatványszerűen lecsengő, gyenge korreláció, míg $\alpha \geq 1$ mellett a divergáló, vagyis erős korreláció esete vizsgálható. Mivel az $1/f^\alpha$ jelek a Fourier-térben független Gauss-eloszlású módusokból tevődnek össze, ezért szimulációjuk egyszerű és gyors. Az általunk elsősorban vizsgált mennyiség a T hosszúságú periodikus jel maximális magassága volt [32], továbbá annak sorrend statisztikája [33], különös tekintettel az asztrofizika számára is érdekes első és második legnagyobb érték távolságára.

A legnagyobb érték vizsgálata során a numerikus és analitikus eredményeink [32] azt mutatták, hogy az $\alpha = 1$ paraméter két lényegesen különböző tartományt választ el a $T \rightarrow \infty$ esetén kapható határeloszlást és az ahhoz való konvergenciát illetően. Gyenge korreláció ($\alpha < 1$) esetén az iid esettel megegyező határeloszlást kapunk, azaz az eloszlások az FTG-hez tartanak végtelen mintaméretre. A szimulációk szerint a véges méret korrekció függvényének alakja sem változik $\alpha < 1$ esetén az iid esethez képest. A konvergencia azonban $\alpha = 0.5$ -nél megváltozik: $\alpha < 0.5$ -ig a korrelálatlan esettel megegyező (logaritmikus), az $0.5 < \alpha < 1$ tartományban ezzel szemben egyre lassul a növekvő korrelációval. Erős korreláció jelenléte ($\alpha \geq 1$) esetén a konvergencia gyors, de a $T \rightarrow \infty$ határeloszlás az FTG-től különböző, analitikus alakja egyelőre csak néhány partikuláris esetben ismert, úgymint $\alpha = 1$ [44], $\alpha = 2$ [30] és $\alpha = \infty$ [32].

A korreláció hatását vizsgáltuk $1/f^\alpha$ típusú zajok sorrend statisztikájában is [33], ahol a rendezett minta elemei között várható távolságok skálázását néztük, és annak változását a korreláció mértékével. A sorrend statisztikán belül az alkalmazások szempontjából egy érdekes mennyiség az első és második legnagyobb érték várható távolsága. Az asztrofizikában például a klaszter galaxisok fényessége kapcsán került az érdeklődés középpontjába ez a mennyiség, mivel a vártnál nagyobb értéke a klaszter legfényesebb galaxisának eltérő statisztikáját sugallta [65, 67]. A távolság $N \rightarrow \infty$ határértéke iid változókra az EVS univerzális osztályaira jellemző érték, és a három osztályra – γ -val paraméterezve – egységes formulával adható meg. Bár a távolság határértéke egyszerű módon származtatható a határeloszlások segítségével, annak véges méret korrekciója eddig ismeretlen volt, ezért az asztrofizikai alkalmazás ér-

dekében elsőrendű tagját egzaktul meghatározzuk a Fisher-Tippett-Gumbel osztályra. Dolgozatunk második fő témakörét adja a korrelált változók sorrend statisztikájának leírása $1/f^\alpha$ jelek vizsgálatával, és az első-második legnagyobb közötti távolság vizsgálata iid változókra.

Az EVS alkalmazása során a vizsgálni kívánt adatsort azonos méretű cellákra szoktuk osztani, és a maximumot ezeken a cellákon értelmezve keressük. Előfordulhatnak azonban olyan helyzetek, amikor – az adatsor hiányossága vagy a változó nem egyenletes eloszlása miatt – nem lehetséges azonos méretű cellák kialakítása a statisztikai analízishez használt empirikus adatbázisban. Más szóval ilyenkor a maximumokat olyan cellákból figyeljük meg, amelyek mérete, N , egy véletlen változó, és celláról cellára különböző értéket vesz fel. Jogosan merül fel a kérdés, hogy ezekből a maximumokból összeállított histogram mennyiben tér el a nem véletlenszerű, determinisztikus N mintaméret mellett várható extrém eloszlástól. A 'klasszikus' értelemben vett EVS csak determinisztikus mintaméret mellett írja le a maximumok statisztikáját.

Általános esetben, tetszőleges mintaméret eloszlás és őseloszlás mellett, vizsgáljuk az extrém statisztikát, és azt látjuk, hogy a mintaméret eloszlás skálázási tulajdonságaitól függ az extrém határeloszlás. A végtelen mintaméret határesetét – amelyben a határeloszlás megjelenik – ilyenkor úgy értelmezzük, hogy az N -eloszlás egy tipikus értékével, N_0 -lal tartunk végtelenhez. Nincs hatása az extrém határeloszlásokra az N véletlenszerűségének, ha a határátmenet során a tipikus N_0 értékkel átskálázott N -eloszlás összeeszkül, ilyenkor a határeloszlások megegyeznek a fix N esettel. Ha azonban az N -eloszlás skálázik, a határeloszlás megváltozik, és függ az N -eloszlás skálafüggvényétől. Az eredeti extrém elmélettel azonos vonás, hogy adott N -eloszlás mellett, az őseloszlás aszimptotikája szerint három univerzális határeloszlás család jelenik meg. Bár az EVS problémája véletlen mintaméret mellett megjelent a matematikai irodalomban és formálisan ismert, hogy adott N -eloszlás mellett az extrém határeloszlás zárt formában meghatározható [46], használata az alkalmazási területeken egyelőre nem elterjedt. A véges méret korrekció azonban eddig ismeretlen maradt, amit skálázó mintaméret eloszlás mellett vizsgálunk meg és azt találjuk, hogy a fix N EVS kapcsán bevezetett RG megközelítés kiterjeszthető a véletlen mintaméretre is, a marginális esetben. Dolgozatunk harmadik fő témaköre, a fentiekben leírt, változó mintaméret problematikája.

Az extrém statisztika eleve az adatok egy kis hányadával dolgozik – mivel N adatból csak egyet, a maximumot használja –, éppen ezért csak elegendően nagy adatbázisok esetén nyílik lehetőség a vizsgálatára. Az elmúlt évek kiterjedt galaxis katalógusai (mint a Sloan Digital Sky Survey - SDSS) nagyszámú adattal szolgálnak ahhoz, hogy az asztrofizika területén is felmerüljön

az extrém statisztika kérdése [48, 69]. Az EVS fenti három témakörében kapott eredményeinket galaxis fényesség értékek extrém statisztikai vizsgálatában használjuk fel. A megfigyelt mennyiség egy irány adott térszögű környezetében látható legfényesebb galaxis luminozitása, ennek az empirikus eloszlását vizsgáljuk elméleti eredményekkel összevetve és szimulációkkal egyaránt. Az extrém statisztika vizsgálatához a teljes adat bázist szín, morfológia és eloszlás szerint négy fő populációra osztjuk, viszont azt tapasztaljuk hogy a különböző populációk galaxiszám eloszlásai egy paraméterrel összekálázhatók. Az adathalmaz kialakításához az égboltot egyenlő kiterjedésű térszögekre (pencil beam geometriájú térfogategységekre) osztjuk, ezért a mintavételi cellák, amelyből a maximumot megfigyeljük, eltérő számú galaxist tartalmaznak. Ugyanakkor egy adott térszögben megfigyelt galaxisok száma nem túl nagy, emiatt mind a véges mind a véletlen mintaméretből származó eltérésekre kapott eredményeink hasznosnak bizonyulnak az empirikus eloszlások értelmezésében. Az őseloszlás, a luminozitás eloszlás, alapján a maximális galaxis fényességek határeloszlása a Fisher-Tippett-Gumbel-eloszlás, és bár a galaxis fényességek korrelált változók, az eloszlások az iid elmélet eredményeivel jól magyarázhatók. Megvizsgáljuk továbbá az első és második legfényesebb galaxis magnitúdó-gap értékét és a Tremaine-Richt one hányadosokat [65] véges méretű mintában. A dolgozat negyedik témaköre a galaxis fényességek extrém statisztikája köré épül. Minden fejezet a továbbiakban egy bevezetővel indul, amelyben röviden össze vannak foglalva az eredmények, s ezután kerül sor azok részletezésére.

Amennyiben megfelelően nagy adathalmaz áll rendelkezésre az extrém érték statisztika a fenti asztrofizikai alkalmazáshoz hasonlóan számos területen használható. A véges és véletlen mintaméret problémáinak részletes vizsgálatával az empirikus extrém eloszlások pontosabb leírását céloztuk meg, az EVS aszimptotikus elméletétől való eltérések értelmezése által. Úgy gondoljuk, hogy ezek az eltérések fontosak, elég általánosak, s az általunk kapott eredmények hasznosak lesznek későbbi extrém statisztikai analízisekben is.

2. fejezet

EXTRÉM ÉRTÉK STATISZTIKA RENORMÁLÁSI CSOPORT MEGKÖZELÍTÉSE

Az extrém érték statisztika aszimptotikus elmélete iid változókra relatíve egyszerű. Ekkor a minta méretével a végtelenhez tartva univerzális határeloszlások jelennek meg. Kevesebbet tudunk azonban a véges mintaméret esetéről. A klasszikus leírás átfogalmazható egy renormálási csoport transzformáció bevezetésével, ebben a megközelítésben a már ismert eredmények [1, 2] egyszerűen és könnyen átláthatóan kaphatók meg. Az új megközelítésnek a legnagyobb érdeme, hogy a határeloszlásokon túl meghatározható a véges méretből adódó korrekció függvénye.

Az EVS határeloszlásai kapcsán felmerülő véges méret viselkedést a matematikai irodalom leírta iid változókra [23, 25], megadva a konvergencia sebességét és a korrekció függvényét. Ez az eredmény a fizikai irodalom számára azonban sokáig ismeretlen volt, és az alkalmazások között sem terjedt el. Nemrégiben vizsgálták meg korrelált változók egy speciális esetét [34], és azt találták, hogy a véges méret korrekció a határeloszlás deriváltjaként adódik. A terület legújabb eredménye, hogy egy új – renormálási csoport (RG) – megközelítés bevezetésével a véges méret korrekció eredményei könnyen kezelhető és érthető formában kerültek általunk bemutatásra, és a korrekció függvénye leírásra került iid változókra, az alkalmazási területeken leggyakrabban használt sztenderdizációban [26, 27]. Mivel a határeloszlásokhoz való konvergencia általában lassú, és az alkalmazások során a vizsgálni kívánt idősorok hossza korlátozott, ezért a véges méret korrekció ismerete lényeges, annak érdekében hogy az ilyen idősorokban megjelenő maximum statisztikáját kellő pontossággal tudjuk leírni. A lassú konvergenciát jól illusztrálja a

Fisher-Tippett-Gumbel eloszlás, amihez a legtöbb őseloszlásból kiindulva a konvergencia logaritmikus, $1/\ln N$ -nel arányos a mintaméret növelése során.

A statisztikus fizikában a kritikus jelenségek terén bevált módszer a renormálási csoport, mivel lehetővé teszi egy szinguláris jelenség véges transzformációval való leírását. Az itt vázolt RG kép alap gondolata nem más, mint hogy megadja az EVS $N \rightarrow \infty$ limesz problémáját azáltal, hogy minden véges N -re egy transzformációt definiál. A már ismert invariancia feltételt [2] a transzformáció fixpont egyenleteként fogjuk fel, melynek megoldása a – határeloszlásokat egy formulába összesítő – γ -val parametrizált egyesített extrém határeloszlás, ami az RG képben a transzformáció fix vonalaként áll elő. A transzformáció hatása a fixpont környezetében a méret változását eredményezi, a vonzó sajátirányok mentén a határeloszláshoz, mint fixponthoz való konvergenciát. A fixpont körüli aszimptotikus viselkedést a domináns sajátérték határozza meg, a véges méret korrekció vezető rendben ezért felírható a transzformáció sajátfüggvénye segítségével, a konvergencia sebessége pedig a sajátértékkel kapcsolatos. Az FSC kapcsán – a határeloszlásokhoz hasonlóan – univerzális vonások jelennek meg, ugyanis egy adott γ -jú határeloszlás esetén a véges méret korrekciót egy extra paraméterrel, γ' -vel, jellemezhetjük.

Ebben a fejezetben összefoglalást adunk az EVS elméletének alapelemeiről, bemutatjuk az RG megközelítés főbb eredményeit [26, 27] alapján, valamint az FSC függvényének meghatározását adott őseloszlású iid változók esetére, és néhány, az alkalmazások szempontjából érdekes őseloszlás esetét külön részletezzük.

2.1. Bevezetés az iid változók extrém érték statisztikájába

Az extrém érték statisztika (EVS) véletlen változók extrém értékének eloszlását, statisztikai tulajdonságait írja le. A vizsgálat tárgya egy N elemszámú véletlen változókból álló minta z_1, z_2, \dots, z_N , a kérdés ezek legnagyobb értékének $z = \max\{z_1, z_2, \dots, z_N\}$ eloszlása (a minimum eloszlása egy előjelváltás után ezzel ekvivalens probléma). A válasz jól ismert abban a klasszikus esetben, ha az $N \rightarrow \infty$ limitet vesszük, valamint z_1, z_2, \dots, z_N véletlen változó azonos eloszlású és egymástól független (iid) [1, 2].

Ekkor az őseloszlás (a z_1, z_2, \dots, z_N változók eloszlása) aszimptotikus viselkedése választja ki, hogy a három extrém határeloszlás-család – Fisher-Tippett-Gumbel (FTG), Fisher-Tippett-Fréchet (FTF), vagy Fisher-Tippett-Weibull (FTW) – közül melyikhez fog tartani a maximumok eloszlása az

$N \rightarrow \infty$ limitben. Ha az őseloszlás minden hatványnál gyorsabban cseng le (akár véges értéknél akár a végtelenben) a Fisher-Tippett-Gumbel-eloszlás vonzási tartományába tartozik, ha lecsengése a végtelenben hatványszerű a Fisher-Tippett-Fréchet-eloszlás, illetve ha az őseloszlás véges értéknél egy hatványkitevő szerint tart 0-hoz a Fisher-Tippett-Weibull-eloszlás vonzási tartományába.

A fentieket precízebben megfogalmazva, jelölje $\rho(z)$ az őseloszlást (parent distribution), vagyis az eredeti változó, z_1, z_2, \dots, z_N eloszlásának sűrűségfüggvényét, míg

$$\mu(z) = \int_{-\infty}^z \rho(t) dt \quad (2.1.1)$$

az integrált eloszlásfüggvényét, ami annak a valószínűségét számszerűsíti, hogy a véletlen változó kisebb, mint z . Ennek használatával könnyen felírható az extrém eloszlás. Annak a valószínűsége, hogy az N elemű minta maximumának értéke kisebb, mint z , értelemszerűen nem más, mint annak a valószínűsége, hogy az összes változó kisebb, mint z , vagyis az N elemű minta extrém érték eloszlása (EVD) a változók függetlensége esetén

$$M_N^{\max}(z) = \mu(z)^N, \quad (2.1.2)$$

és az extrém eloszlás sűrűségfüggvénye

$$P_N^{\max}(z) = dM_N^{\max}(z)/dz = N\mu(z)^{N-1}\rho(z). \quad (2.1.3)$$

Mivel $\mu(z)$ integrált eloszlásfüggvény, tehát értéktartománya 0-tól 1-ig tart, $M_N^{\max}(z)$ minden pontja 0-hoz tart a tartó belsejében az $N \rightarrow \infty$ határátmenetben. Az EVS elmélete szerint létezik olyan lineáris transzformáció, amellyel z -t átskálázva ($z = a_N x + b_N$) az $N \rightarrow \infty$ limitben egy nem degenerált határeloszlást kaphatunk, vagyis

$$M_N(x) = M_N^{\max}(a_N x + b_N) \rightarrow M(x). \quad (2.1.4)$$

Az extrém eloszlás sűrűségfüggvényére $P_N(x) = dM_N(x)/dx$ megfogalmazva

$$P_N(x) = a_N P_N^{\max}(a_N x + b_N) \rightarrow P(x), \quad (2.1.5)$$

ahol $P(x) = M'(x)$ az extrém határeloszlás.

Keressük tehát azokat az a_N, b_N együtthatókat, amellyel a változót áttranszformálva a fenti limesz megvalósul. Mivel a_N és b_N bizonytalan egy additív konstans erejéig, további sztenderdizációval fixáljuk azokat (l. később (2.5.1) és (2.5.2)).

Az extrém értékek elméletének alapvető eleme az univerzalitási osztályok megjelenése az $N \rightarrow \infty$ limeszben. Ha a limesz létezik, az extrém eloszlás a három határeloszlás család egyikéhez fog tartani, ha $N \rightarrow \infty$, amelyek egyégesen leírhatók egy formulával, ez az egyesített extrém határeloszlás függvény (GEV)

$$M(x; \gamma) = e^{-(1+\gamma x)^{-1/\gamma}}, \quad (2.1.6)$$

$$P(x; \gamma) = (1 + \gamma x)^{-1/\gamma-1} e^{-(1+\gamma x)^{-1/\gamma}}, \quad (2.1.7)$$

ahol $1 + \gamma x \geq 0$, és γ értéke szerint különböztetjük meg a határeloszlás családokat. A $\gamma = 0$ érték a Fisher-Tippett-Gumbel eloszlásnak felel meg

$$M(x; 0) = e^{-e^{-x}}, \quad (2.1.8)$$

$$P(x; 0) = e^{-x-e^{-x}}, \quad (2.1.9)$$

míg $\gamma < 0$ a Fisher-Tippett-Weibull, illetve $\gamma > 0$ a Fisher-Tippett-Fréchet eloszlás családot adja meg. γ értékét, vagyis azt, hogy melyik határeloszláshoz tart az extrémumok eloszlása, az őseloszlás aszimptotikus viselkedése határozza meg. Az a_N és b_N paraméterek határozatlansága miatt a határeloszlások a fenti formuláktól egy lineáris transzformáció erejéig eltérhetnek.

2.2. A renormálási csoport transzformáció fixpontjai és az extrém határeloszlások

A renormálási csoport megközelítés alapja a matematikai irodalom által max-stability néven ismert invariancia feltétel [2]. Eszerint ha $M(x)$ extrém határeloszlás, és ebből mint őseloszlásból húzunk p darab változót, azok maximumának eloszlása ugyanez az $M(x)$ eloszlás lesz egy lineáris transzformáció erejéig, formálisan

$$M(x) = M^p[a(p)x + b(p)]. \quad (2.2.1)$$

Ez az állítás jól látszik akkor, ha ezeket a változókat egy nagyobb adathalmaz cella maximumainak tekintjük, ezért a maximumok maximuma is nyilván ugyanazt a határeloszlást fogja követni. Bár p itt pozitív valós, a feltétel kiterjeszthető pozitív racionális számokra. Az $a(p), b(p)$ paramétereket a sztenderdizáció határozza meg, a továbbiakban az alábbi feltételek teljesítésével

$$M(0) = P(0) = 1/e. \quad (2.2.2)$$

Az EVS renormálási csoport megközelítése [26, 27] alapján az RG transzformációt egy $\mu(x)$ integrált eloszláson – (2.2.1) jobb oldala alapján – a következő operátor definiálja

$$[\hat{R}_p\mu](x) = \mu^p[a(p)x + b(p)]. \quad (2.2.3)$$

A (2.2.2) sztenderdizációt alkalmazva az $[\hat{R}_p\mu](x)$ transzformált eloszlásra, meghatározhatjuk az $a(p), b(p)$ paramétereket

$$b(p) = \mu^{-1}(e^{-1/p}), \quad (2.2.4a)$$

$$a(p) = p \frac{db(p)}{dp}. \quad (2.2.4b)$$

Ha $M(x)$ extrém határeloszlás, akkor tetszőleges p -re teljesíti (2.2.1) feltételt, ezért $M(x)$ az RG (2.2.3) szerint definiált transzformációjának fixpontja

$$[\hat{R}_pM](x) = M^p(a(p)x + b(p)) = M(x). \quad (2.2.5)$$

Meghatározhatjuk tehát az EVS határeloszlásokat a fenti fixpont egyenlet megoldásával. A megoldást az

$$M(x) = e^{-e^{-f(x)}} \quad (2.2.6)$$

alakban keressük, ahol a (2.2.2) sztenderdizáció miatt $f(0) = 0$, $f'(0) = 1$ feltételeknek teljesülniük kell. Az $a(p), b(p)$ paraméterek (2.2.4a) és (2.2.4b) alapján f függvényében a következőképpen néznek ki

$$b(p) = f^{-1}(\ln p), \quad (2.2.7a)$$

$$a(p) = \frac{1}{f'(b(p))}. \quad (2.2.7b)$$

A (2.2.6) alakot a (2.2.5) fixpont egyenletbe helyettesítve megoldásként $f = \gamma^{-1} \ln(1 + \gamma x)$ -et kapunk, ami a jól ismert egyesített extrém határeloszlást (GEV) eredményezi [27]

$$M(x; \gamma) = e^{-(1+\gamma x)^{-1/\gamma}}, \quad (2.2.8a)$$

$$P(x; \gamma) = (1 + \gamma x)^{-1/\gamma-1} e^{-(1+\gamma x)^{-1/\gamma}}, \quad (2.2.8b)$$

ahol $1 + \gamma x \geq 0$. A lineáris transzformáció paraméterei pedig

$$a(p, \gamma) = p^\gamma, \quad (2.2.9a)$$

$$b(p, \gamma) = \gamma^{-1}(p^\gamma - 1). \quad (2.2.9b)$$

Az RG megközelítést tekintve az EVS határeloszlások a transzformáció fixpontjai, az általános, γ -val paraméterezett GEV eloszláscsalád pedig egy fixvonal az eloszlások terében, amelyen $\gamma < 0, = 0, > 0$ paraméter értékek a klasszikus EVS univerzalitási osztályainak, nevezetesen az FTW, FTG illetve FTF eloszlásnak felelnek meg.

2.3. Perturbációk a fixpont körül és a véges méret skálázás

Az EVS renormálási csoport megközelítésének igazi jelentősége a véges méret korrekció leírásában rejlik. Az FSC függvénye ugyanis megkapható a transzformáció fixpont körüli linearizálásából.

Az RG transzformációja egy N méretű minta maximumának eloszlását egy $N' = pN$ méretű minta maximumának eloszlásába transzformálja

$$[\hat{R}_p M_N](x) = M_N^p(a(p, N)x + b(p, N)) = M_{N'}(x), \quad (2.3.1)$$

mert $M_N(x) = \mu^N(a_N x + b_N)$. Vagyis a transzformáció hatása egy extrém eloszláson nem más, mint a minta méretének növelése (ha $p > 1$), a véges méretű extrém eloszlás $M_N(x)$ közelebb kerül az $M(x) = \lim_{N \rightarrow \infty} M_N(x)$ határeloszláshoz, ami a transzformáció egy vonzó fixpontja.

Az aszimptotikus véges méret viselkedés ezért az RG transzformációjának a fixpont környékén gyakorolt hatásával vizsgálható [26]. A véges méretű extrém eloszlást a fixpont körüli perturbáció formájában keressük

$$M(x; \gamma, \epsilon) = M(x + \epsilon \psi(x); \gamma), \quad (2.3.2)$$

ahol $M(x; \gamma)$ a fixpont, és ϵ kicsi. Lineáris rendig megegyezően ez a perturbált eloszlásfüggvény az

$$M(x; \gamma, \epsilon) = M(x; \gamma) + \epsilon P(x; \gamma) \psi(x) \quad (2.3.3)$$

alakban is írható. Mivel a perturbált eloszlásfüggvényt is (2.2.2) alapján kell sztenderdizálni, ez a következő feltételeket szabja a $\psi(x)$ függvényre

$$\psi(0) = 0, \quad \psi'(0) = 0, \quad \psi''(0) = -1, \quad (2.3.4)$$

ahol a harmadik megkötéssel az ϵ skáláját rögzítjük. Ha feltételezzük, hogy az ilyen módon perturbált eloszlásfüggvény formailag nem változik az RG transzformáció hatására, csak az ϵ paraméter értéke változik ϵ' -re

$$[\hat{R}_p M(\epsilon)](x) = M^p(a(p, \epsilon)x + b(p, \epsilon); \epsilon) = M(x; \epsilon'), \quad (2.3.5)$$

akkor látjuk, hogy az $M(x, \epsilon)$ valóban jó közelítése a véges méretű extrém eloszlásnak, hiszen a (2.3.1) által leírt véges méret viselkedést kaptuk vissza, ahol az ϵ tartalmazza az N -függést. (2.3.3)-at behelyettesítve látszik, hogy $\psi(x)$ és $\lambda = \epsilon'/\epsilon$ egy sajátérték probléma megoldásaként áll elő, mint sajátfüggvény és ahhoz tartozó sajátérték. A megoldás során egy új paraméter,

γ' , jelenik meg, amely a sajátfüggvényeket jellemzi. A sajátérték probléma általános megoldása így tehát két paramétert tartalmaz [27]

$$\psi(x; \gamma, \gamma') = \frac{1 + (\gamma' + \gamma)x - (1 + \gamma x)^{\gamma'/\gamma+1}}{\gamma'(\gamma' + \gamma)}, \quad (2.3.6)$$

$$\lambda(\gamma') = p^{\gamma'}. \quad (2.3.7)$$

Ez az elsőrendű korrekciót meghatározó függvényalak ekvivalens a matematikai irodalom konvergenciát vizsgáló korábbi eredményeivel [23, 24]. Az FTG osztályra, amit $\gamma = 0$ paraméterez a fix vonalon, a $\psi(x)$ sajátfüggvényt a fenti formula $\gamma \rightarrow 0$ határértékeként kapjuk meg

$$\psi(x; 0, \gamma') = \frac{1 + \gamma'x - e^{\gamma'x}}{\gamma'^2}. \quad (2.3.8)$$

Fontos szerepet játszik az a speciális eset, amikor $\gamma = \gamma'$, ekkor a sajátfüggvény a következő γ -tól független alakra egyszerűsödik

$$\psi(x; \gamma, \gamma) = -\frac{x^2}{2}. \quad (2.3.9)$$

Az új paraméter, γ' , gyakorlati jelentősége, hogy összekapcsolja a perturbáció kis paraméterét ϵ -t a véges méretet jellemző N -nel. (2.3.1) és (2.3.5) összevetésével láthattuk, hogy az N -függést az ϵ írja le, ezért a továbbiakban ϵ_N -ként használjuk, továbbá leolvashatjuk, hogy

$$\epsilon' = \epsilon_{pN} = \lambda(\gamma')\epsilon_N. \quad (2.3.10)$$

Ekkor (2.3.7)-et felhasználva és ϵ_N -ben hatványfüggést feltételezve kapjuk, hogy

$$\epsilon_N \propto N^{\gamma'}. \quad (2.3.11)$$

Míg γ a határeloszlást meghatározó paraméter, addig γ' a véges méret skálázást jellemzi, a konvergencia sebességét leíró exponens, ahol $\gamma' < 0$ a stabilitás érdekében. A $\gamma' = 0$ a stabilitást illetően marginális eset, vagyis lineáris közelítésben az RG transzformáció nem viszi közelebb a véges méretű eloszlást a határeloszláshoz. Az FTG ($\gamma = 0$) osztályon belül azonban a marginális eset a tipikus, amint ezt látni is fogjuk a 2.6. fejezetben, ebben a megközelítésben az ϵ_N hatványnál lassabban lecsengő N függését jelenti, ez magyarázza az FTG osztályban olyan gyakran megjelenő logaritmikusan lassú konvergenciát.

Az őseloszlás aszimptotikus viselkedése határozza meg a γ és γ' értékét, γ paraméter a határeloszlást választja ki a fix vonalon, γ' pedig az adott

határeloszláshoz (γ -hoz) tartozó FSC sajátfüggvényét (2.3.6) és annak amplitúdóját megadó ϵ_N -t, ezzel együtt a véges méretű extrém eloszlás első rendig pontosan megkapható. Véges méretű minta esetén tehát az extrém eloszlás véges méretben a következőképpen írható fel

$$M_N(x) = M(x; \gamma) + \epsilon_N M_1(x; \gamma, \gamma') = M(x; \gamma) + \epsilon_N P(x; \gamma) \psi(x; \gamma, \gamma'), \quad (2.3.12)$$

valamint deriválással a sűrűségfüggvénye

$$\begin{aligned} P_N(x) &= P(x; \gamma) + \epsilon_N P_1(x; \gamma, \gamma') = \\ &= P(x; \gamma) + \epsilon_N P(x; \gamma) \left[\frac{P'(x; \gamma)}{P(x; \gamma)} \psi(x; \gamma, \gamma') + \psi'(x; \gamma, \gamma') \right], \end{aligned} \quad (2.3.13)$$

és $\psi(x)$ a (2.3.6) általános formulával meghatározott sajátfüggvény. A későbbiekben az egyszerűbb leírás érdekében $M_1(x)$ illetve $P_1(x)$ jelöli az elsőrendű véges méret korrekció függvényét.

2.4. A véges méret korrekció és az őseloszlás kapcsolata

Az elsőrendű véges méret korrekció az RG megközelítés alapján felírható (2.3.12) szerint, ha az őseloszlás vizsgálatával meghatározzuk a szükséges paramétereket $(\gamma, \gamma', \epsilon_N)$. Ezen kérdéses paraméterek értéke csak az őseloszlás aszimptotikus viselkedésétől függ.

A paraméterek meghatározásához az integrált őseloszlást

$$\mu(z) = e^{-e^{-g(z)}} \quad (2.4.1)$$

alakban vizsgáljuk. A $z = a_N x + b_N$ lineáris transzformációval sztenderdizált extrém eloszlás (2.1.4)-ben ekkor

$$M_N(x) = \mu(a_N x + b_N)^N = e^{-N e^{-g(a_N x + b_N)}}, \quad (2.4.2)$$

ahol $M_N(x) \rightarrow M(x)$, és az a_N, b_N paramétereket a (2.2.2) feltételek határozzák meg, miszerint

$$g(b_N) = \ln N, \quad (2.4.3)$$

$$g'(b_N) = \frac{1}{a_N}. \quad (2.4.4)$$

Ugyanakkor feltesszük, hogy elég nagy N esetén a véges méretű extrém eloszlás egy sajátírány mentén konvergál a fixponthoz, (2.3.2) szerint

$$M_N(x) \approx M\left[x + \epsilon_N \psi(x; \gamma, \gamma'); \gamma\right], \quad (2.4.5)$$

ahol (2.2.6) és (2.4.2) helyettesítés után, kétszer negatív logaritmust véve és (2.4.3)-at kihasználva

$$g(a_N x + b_N) - g(b_N) \approx f(x; \gamma) + \epsilon_N f'(x; \gamma) \psi(x; \gamma, \gamma'). \quad (2.4.6)$$

Ezt kétszer deriválva $x = 0$ -ban és a sztenderdizáció által kirótt feltételeket, valamint $f''(0; \gamma) = -\gamma$ tulajdonságot és (2.4.4)-et kihasználva, az ϵ_N -t a következőképpen fejezhetjük ki

$$\epsilon_N = \gamma_N - \gamma, \quad (2.4.7)$$

ahol

$$\gamma_N = -\frac{g''(b_N)}{g'^2(b_N)} = \frac{da_N}{db_N} \rightarrow \gamma. \quad (2.4.8)$$

Ezek alapján az őseloszlás ismeretében a következőképpen határozható meg a véges méret korrekció. (2.4.3) és (2.4.4) formulák megadják a sztenderdizáció paramétereit, ahol $g(z) = -\ln(-\ln \mu(z))$ függvényt elegendő első rendig sorbafejteni nagy z értékekre. Az a_N, b_N ismeretében megkapjuk γ_N -t (2.4.8), valamint a határeloszlást meghatározó γ exponenst a limesz határértékeként az $N \rightarrow \infty$ limitben. A véges méret korrekció amplitúdóját (2.4.7) különbség adja meg, ahonnt, mivel $\epsilon_N \propto N^{\gamma'}$, megkapható γ' , a véges méret exponense, ami meghatározza a véges méret korrekció függvényében szereplő $\psi(x; \gamma, \gamma')$ sajátfüggvényt [27].

2.5. Az eloszlások és a véges méret korrekció alakja a két sztenderdizációban

Ahogy azt a korábbiakban láthattuk, az extrém határeloszláshoz való konvergencia során a véges extrém eloszlásokat egy lineáris transzformációval sztenderdizáljuk (2.1.5), hogy a divergenciát kiküszöböljük. A transzformáció paraméterei azonban bizonytalanok egy additív konstans erejéig, ezért végtelen sok ilyen transzformáció létezik. A következő fejezetek során a (2.2.2) sztenderdizáció mellett egy másikat is fogunk használni, a kétféle sztenderdizációt (1) illetve (2) jelöli az alábbi feltételek szerint.

Az eddigiekben is használt (2.2.2) feltételeket fogjuk (1) típusú sztender-

dizációnak nevezni, vagyis

$$M(x=0) = 1/e \quad P(x=0) = 1/e. \quad (2.5.1)$$

Az eloszlás átlagának és szórásának rögzítésével definiáljuk a (2) típusú szten-derdizációt

$$\int_{-\infty}^{\infty} xP(x)dx = 0 \quad \int_{-\infty}^{\infty} x^2P(x)dx = 1. \quad (2.5.2)$$

Az analitikus megfontolások szempontjából érdemes az (1) típusú szten-derdizációt használni, mert a határeloszlások alakja ebben a legegyszerűbb, a gyakorlati alkalmazások során viszont sokszor a (2) típusú szten-derdizáció használata kézenfekvőbb, természetesen amennyiben $P(x)$ eloszlás szórása létezik. Az empirikus, mindig véges méretű eloszlásokat legegyszerűbben úgy tudjuk összevetni egymással illetve az elméleti határeloszlással, ha azokat az $(x - \langle x \rangle)/\sigma_x$ lineáris transzformációval egységnyi szórásra és nulla várható értékűre skálázzuk. A határeloszlásokat $(P(x), M(x))$, vagy az elméleti véges méretű eloszlásokat $(P_N(x), M_N(x))$ átskálázni a (2) típusú szten-derdizációba egyszerű, egy $ax + b$ változócserevel végrehajtható, ahol a az eloszlás szórása b a várható értéke. A véges méret korrekció átszten-derdizálása nem ennyire triviális, különbségfüggvény lévén $(\epsilon_N P_1(x) = P_N(x) - P(x))$, ahol $P_N(x)$ és $P(x)$ függvényében is változócsere történik – más-más paraméterekkel – vagyis a két függvény egymáshoz képest eltolódik, és a különbségfüggvény alakja megváltozik. A véges méretű extrém eloszlás paraméterei (2.3.13) szerint a következőképpen sorbafejtethők

$$b_N = \langle x \rangle_{P_N(x)} = \int_{-\infty}^{\infty} xP_N(x)dx = b + \epsilon_N b_1, \quad (2.5.3)$$

$$a_N^2 = \sigma_{P_N(x)}^2 = \int_{-\infty}^{\infty} x^2P_N(x)dx - b_N^2, \quad (2.5.4)$$

$$a_N = a + \epsilon_N a_1,$$

ahol a paraméterek

$$b = \langle x \rangle_{P(x)}, \quad a = \sigma_{P(x)}, \quad b_1 = \langle x \rangle_{P_1(x)}, \quad a_1 = \frac{\langle x^2 \rangle_{P_1(x)} - 2bb_1}{2a} \quad (2.5.5)$$

Az (1)→(2) átszten-derdizálás nem más, mint az

$$x \longrightarrow a_N x + b_N = ax + b + \epsilon_N(a_1 x + b_1) \quad (2.5.6)$$

változócsere. (2.3.12)-ben végrehajtva ezt a változócsere-t, majd sorbafejtve megkapjuk a véges méretű eloszlást a (2) vagyis a variancia szerinti szten-derdizációban

$$M_N^{\text{var}}(x) = M_N(a_N x + b_N)$$

$$M_N^{\text{var}}(x) = M(ax + b) + \epsilon_N P(ax + b) [\psi(ax + b) + a_1 x + b_1], \quad (2.5.7)$$

vagyis a véges méret korrekció függvénye átsztenderdizálva

$$M_1^{\text{var}}(x) = P(ax + b) [\psi(ax + b) + a_1 x + b_1]. \quad (2.5.8)$$

$M_N^{\text{var}}(x)$ deriválásával – vagy (2.3.13)-ban vett változócsere és sorbafejtés után – kapjuk a sűrűségfüggvényt a (2) sztenderdizációban

$$P_N^{\text{var}}(x) = a_N P_N(a_N x + b_N) = aP(ax + b) +$$

$$+ \epsilon_N aP(ax + b) \left[\frac{P'(ax + b)}{P(ax + b)} (\psi(ax + b) + a_1 x + b_1) + \psi'(ax + b) + \frac{a_1}{a} \right], \quad (2.5.9)$$

a sűrűségfüggvény véges méret korrekciójának függvénye

$$P_1^{\text{var}}(x) = aP(ax + b) \left[\frac{P'(ax + b)}{P(ax + b)} (\psi(ax + b) + a_1 x + b_1) + \psi'(ax + b) + \frac{a_1}{a} \right]. \quad (2.5.10)$$

(2.3.12)-t és (2.5.7)-t összevetve látjuk, hogy az (1) \rightarrow (2) átskálázás formálisan a következőképpen írható

$$x \rightarrow ax + b,$$

$$\psi \rightarrow \psi + a_1 x + b_1. \quad (2.5.11)$$

Vegyük észre, hogy a gondolatmenet során sehol sem használtuk ki az (1) sztenderdizáció tulajdonságait, ezért a fenti módszer alkalmas más módon sztenderdizált eloszlás és korrekció függvény (2) sztenderdizációba való átskálázására is.

2.6. Véges méret korrekció az FTG osztályban

Az alkalmazások szempontjából egyik legfontosabb és leggyakrabban használt eset a Fisher-Tippett-Gumbel-eloszlás vonzási tartománya ($\gamma = 0$). A disszertáció 5. fejezetében tárgyalt maximális galaxis luminozitások eloszlása is ide tartozik, ezért érdemes itt ennek az osztálynak a véges méret korrekciójával részletesebben foglalkozni.

Az FTG vonzási tartományába a minden hatványnál gyorsabban lecsengő őseloszlások tartoznak. A minél szélesebb körű leíráshoz tekintsük a következő aszimptotikájú eloszlást, amit általánosított exponenciális őseloszlásnak

fogunk nevezni

$$\mu(z) \approx 1 - \frac{Be^{-Az^\delta}}{z^\theta}, \quad (2.6.1)$$

$z \rightarrow \infty$ -re, és $\delta > 0$. Innét

$$g(z) = -\ln(-\ln \mu(z)) \approx Az^\delta + \theta \ln z - \ln B. \quad (2.6.2)$$

(2.4.3) és (2.4.4) alapján a sztenderdizáció paraméterei vezető rendben

$$b_N = g^{-1}(\ln N) \approx \left(\frac{\ln N}{A} \right)^{1/\delta}, \quad (2.6.3)$$

$$a_N = \frac{1}{g'(b_N)} \approx \frac{b_N^{1-\delta}}{A\delta} - \frac{\theta b_N^{1-2\delta}}{A^2\delta^2}. \quad (2.6.4)$$

a_N és b_N ismeretében (2.4.8) szerint meghatározható a határeloszlás γ paramétere, vagyis az, hogy az őseloszlás melyik határeloszlás család vonzási tartományába tartozik

$$\gamma_N = \frac{da_N}{db_N} \approx \frac{1-\delta}{A\delta} b_N^{-\delta} + \frac{\theta(2\delta-1)}{A^2\delta^2} b_N^{-2\delta}, \quad (2.6.5)$$

ami, $N \rightarrow \infty$ határesetben b_N behelyettesítésével

$$\gamma_N \rightarrow \gamma = 0, \quad (2.6.6)$$

a határeloszlás – amint azt az általánosított exponenciális őseloszlás esetén jól tudjuk – valóban az FTG. A határeloszláshoz tartozó véges méret korrekció amplitúdója (2.4.7) alapján vezető rendben

$$\epsilon_N = \gamma_N \approx \begin{cases} \frac{1-\delta}{\delta \ln N}, & \delta \neq 1, \\ \frac{\theta}{\ln^2 N}, & \delta = 1, \end{cases} \quad (2.6.7)$$

aminek nagy N értékekre való lecsengése definiálja a γ' véges méret exponens értékét (2.3.11). Itt ϵ_N mindkét esetben logaritmikusan függ N -től, vagyis

$$\gamma' = 0. \quad (2.6.8)$$

A véges méret korrekció vezető rendben (2.3.12) szerint

$$\epsilon_N M_1(x; \gamma, \gamma') = \epsilon_N P(x; \gamma) \psi(x; \gamma, \gamma'), \quad (2.6.9)$$

aminek alakját (2.3.6) sajátfüggvény határozza meg. Az általánosított exponenciális őseloszlásra $\gamma = \gamma' = 0$, vagyis a $\psi(x; \gamma, \gamma')$ sajátfüggvény a (2.3.9) alakra egyszerűsödik. Az FSC függvénye tehát az (1) sztenderdizációban

$$M_1(x; 0, 0) = -\frac{x^2}{2} e^{-x-e^{-x}}, \quad (2.6.10)$$

és a sűrűségfüggvény korrekciója $P_1(x) = M_1'(x)$

$$P_1(x; 0, 0) = e^{-x-e^{-x}} \left[(1 - e^{-x}) \frac{x^2}{2} - x \right]. \quad (2.6.11)$$

Az alkalmazások során gyakran használt (2) sztenderdizációra (2.5.8) illetve (2.5.10) szerint térhetünk át az

$$a = \sqrt{\zeta(2)} = \frac{\pi}{\sqrt{6}}, \quad b = \gamma_E, \quad a_1 = ab + \frac{\zeta(3)}{a}, \quad b_1 = \frac{a^2 + b^2}{2} \quad (2.6.12)$$

paraméterekkel, ahol $\gamma_E = 0.577216$ az Euler-Mascheroni konstans, $\zeta(s)$ a Riemann-féle zeta függvény, s nevezetesen $\zeta(3) = 1.202057$ az ún. Apéry-állandó. A véges méret korrekció függvénye tehát az integrált eloszláshoz a (2) sztenderdizációban

$$M_1^{\text{var}}(x; 0, 0) = e^{-ax-b-e^{-ax-b}} \left(-\frac{a^2}{2}x^2 + \frac{\zeta(3)}{a}x + \frac{a^2}{2} \right), \quad (2.6.13)$$

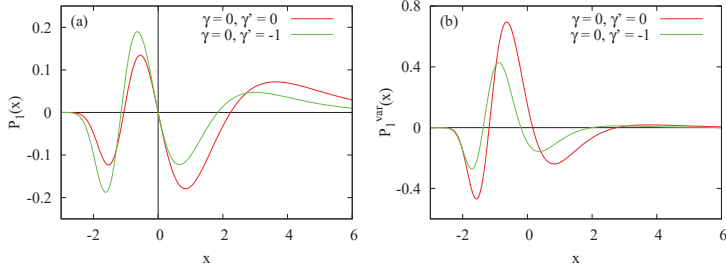
a korrekció függvénye a sűrűségfüggvényhez

$$P_1^{\text{var}}(x; 0, 0) = ae^{-ax-b-e^{-ax-b}} \left[(e^{-ax-b} - 1) \left(-\frac{a^2}{2}x^2 + \frac{\zeta(3)}{a}x + \frac{a^2}{2} \right) - ax + \frac{\zeta(3)}{a^2} \right]. \quad (2.6.14)$$

A logaritmikusan lassú konvergencia egy ismert példája a Gauss-őseloszlás esete ($\rho(x) \propto e^{-x^2}$). Ez az eset a fenti általánosított exponenciálissal leírt aszimptotika $\delta = 2$, $\theta = 1$ értékeinek felel meg. Vagyis (2.6.7) szerint $\epsilon_N = -1/2 \ln N$, ezzel az elsőrendű véges méret korrekció a sűrűségfüggvények szintjén

$$\text{FSC}_{\text{Gauss}} : \quad \epsilon_N P_1(x) = -\frac{1}{2 \ln N} P_1(x; 0, 0), \quad (2.6.15)$$

ahol $P_1(x; 0, 0)$ az FSC függvénye, a használt sztenderdizációtól függően alakját (2.6.11) illetve (2.6.14) adja meg.



2.6.1. ábra. A sűrűségfüggvény véges méret korrekciójának függvényalakja vezet δ rendben, $P_1(x; 0, \gamma') = M_1'(x; 0, \gamma')$ az FTG határeloszláshoz ($\gamma = 0$). Baloldalon az (1), jobb oldalon a (2) szttenderizációban. Az FTG osztályban tipikusan két FSC függvénnyel találkozunk, a $\gamma' = 0$ és -1 -hez tartozóval. Leggyakrabban $\gamma' = 0$ (logaritmusos konvergencia a határeloszláshoz), mint például az általánosított exponenciális δ eloszlás esetén. Kivételt képez a tiszta exponenciális δ eloszlás, amelyre $\gamma' = -1$ (hatványszerű konvergencia).

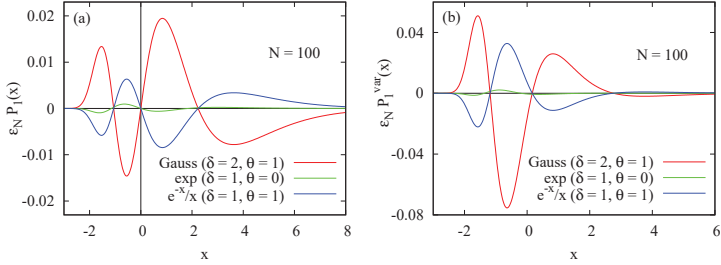
Másik fontos példa a $\delta = 1, \theta = 1$, ami a $\rho(x) \propto e^{-x}/x$ δ eloszlás esete. Ekkor (2.6.7) alapján $\epsilon_N = 1/\ln^2 N$, tehát a véges méret korrekció

$$\text{FSC}_{\text{exp}/x} : \quad \epsilon_N P_1(x) = \frac{1}{\ln^2 N} P_1(x; 0, 0). \quad (2.6.16)$$

A disszertáció 5. fejezetében vizsgált asztrofizikai alkalmazási területen lesz ennek az δ eloszlásnak jelentősége, mivel ilyen alakú a galaxisok luminozitás eloszlása, a Schechter-függvény ($\rho(x) \propto x^\alpha e^{-x}$, ahol $\theta = -\alpha \approx 1$, l. 5.2. fejezet).

Fontos észrevenni, hogy az $1/\ln^2 N$ -tel arányos korrekció csak $\delta = 1$ értékénél jelenik meg. Ha $\delta \neq 1$, akkor az exponenciális mellett szereplő hatványkitevő, θ tetszőleges értékére a konvergencia $1/\ln N$ -es, és a korrekció amplitúdója nem függ θ értékétől. Ez a jelenség lényeges lehet, mint például az alkalmazásunkban felmerülő Schechter-függvénynél is az, mivel az illesztések szerint $\delta \approx 1$, és attól függően, hogy (2.6.7) melyik esetét használjuk, a megjelenő korrekció amplitúdója, sőt előjele is különbözhet. Az empirikus extrém eloszlások helyes leírásában ezért lényeges szerepet játszik az δ eloszlás aszimptotikájának pontos ismerete.

Az általánosított exponenciális δ eloszlásra végzett vizsgálat leírja az FTG osztályba tartozó függvények többségének véges méret korrekcióját.



2.6.2. ábra. A sűrűségfüggvény véges méret korrekciója, $\epsilon_N P_1(x)$ három speciális őseloszlás: Gauss-, e^{-x}/x és tiszta exponenciális eloszlás esetén, (a) ábrán az (1) illetve (b) ábrán a (2) sztenderdizációban. Mindhárom őseloszlás az FTG osztályhoz tartozik, de konvergenciájuk sebessége a határeloszláshoz eltérő, rendre $1/\ln N$, $1/\ln^2 N$ illetve $1/N$, (2.6.15), (2.6.16) ill. (2.6.30) alapján, az FSC nagyságrendi különbsége jól látszik már $N = 100$ -ra. Míg a Gauss és az e^{-x}/x őseloszlás esetén csak ϵ_N előjele és nagysága különbözik, addig a tiszta exponenciális esetén a függvény alakja is eltér.

Külön tárgyalásra szorul azonban egy másik jelentős eset, a tiszta exponenciális őseloszlás. (2.6.1) általánosított exponenciális szerint ez a $\delta = 1$, $\theta = 0$ esetnek felel meg, amire (2.6.7) alapján $\epsilon_N = 0$. Az FSC leírásához tehát szükséges a következő rendfigyelembevétele is.

Tekintsük tehát a tiszta exponenciális őseloszlást

$$\mu(z) = 1 - e^{-Az}, \quad (2.6.17)$$

$$g(z) = -\ln(-\ln \mu(z)) \approx Az - \frac{e^{-Az}}{2}, \quad (2.6.18)$$

a második rendet is figyelembe véve. (2.4.3) és (2.4.4) alapján a sztenderdizáció paraméterei

$$b_N = g^{-1}(\ln N) \approx \frac{\ln N}{A}, \quad (2.6.19)$$

$$a_N = \frac{1}{g'(b_N)} \approx \frac{1}{A} - \frac{e^{-Ab_N}}{2A}. \quad (2.6.20)$$

(2.4.8) és (2.4.7) szerint ez az őseloszlás a $\gamma = 0$, FTG vonzási tartományába

tartozik

$$\gamma_N = \frac{da_N}{db_N} \approx \frac{e^{-Ab_N}}{2} = \frac{1}{2N} \rightarrow \gamma = 0, \quad (2.6.21)$$

az FSC amplitúdója pedig

$$\epsilon_N = \gamma_N = \frac{1}{2N}. \quad (2.6.22)$$

A konvergencia tehát a határeloszláshoz gyors, hatványszerű, a véges méret exponens

$$\gamma' = -1. \quad (2.6.23)$$

Vezető rendben a véges méret korrekció (2.3.12) alapján

$$\epsilon_N M_1(x; \gamma, \gamma') = \epsilon_N P(x; \gamma) \psi(x; \gamma, \gamma'), \quad (2.6.24)$$

ahol most $\gamma = 0$ és $\gamma' = -1$, tehát a sajátfüggvény (2.3.8) alapján $\psi(x; 0, -1) = 1 - x - e^{-x}$, így az FSC függvényalakja az (1) sztenderdizációban

$$M_1(x; 0, -1) = (1 - x - e^{-x})e^{-x-e^{-x}}, \quad (2.6.25)$$

valamint a korrekció függvényalakja a sűrűségfüggvényhez

$$P_1(x; 0, -1) = e^{-x-e^{-x}} [(e^{-x} - 1)(2 - x - e^{-x})]. \quad (2.6.26)$$

Áttérni a (2) típusú sztenderdizációra (2.5.8) és (2.5.10) szerint a következő paraméterekkel tudunk

$$a = \frac{\pi}{\sqrt{6}}, \quad b = \gamma_E, \quad a_1 = a - \frac{1}{a}, \quad b_1 = b, \quad (2.6.27)$$

vagyis az integrált korrekció függvénye a (2) sztenderdizációban

$$M_1^{\text{var}}(x; 0, -1) = e^{-ax-b-e^{-ax-b}} \left(1 - \frac{x}{a} - e^{-ax-b}\right), \quad (2.6.28)$$

a sűrűségfüggvény korrekciójának függvénye pedig

$$P_1^{\text{var}}(x; 0, -1) = ae^{-ax-b-e^{-ax-b}} \left[(e^{-ax-b} - 1) \left(1 - \frac{x}{a} - e^{-ax-b}\right) - \frac{1}{a^2} + e^{-ax-b} \right]. \quad (2.6.29)$$

Tiszta exponenciális őseloszlás esetén a sűrűségfüggvény véges méret korrekciója tehát

$$\text{FSC}_{\text{exp}} : \quad \epsilon_N P_1(x) = \frac{1}{2N} P_1(x; 0, -1), \quad (2.6.30)$$

ahol $P_1(x; 0, -1)$ alakját a két tárgyalt sztenderdizációban (2.6.26) illetve (2.6.29) határozza meg.

A disszertációban szereplő alkalmazás szempontjából egy harmadik típusú őseloszlás extrém eloszlásának konvergenciája is érdekes, ez a lognormális eloszlás

$$\rho(x) = \frac{1}{x\sqrt{2\pi}} e^{-\frac{\ln^2 x}{2}}, \quad x > 0. \quad (2.6.31)$$

A véges méret korrekció meghatározásához az integrált őseloszlás aszimptotikus viselkedése illetve annak elsőrendű közelítése érdekes nagy x értékekre

$$\mu(z) = \frac{1}{\sqrt{2\pi}} \int_0^z \frac{1}{x} e^{-\frac{\ln^2 x}{2}} dx \approx 1 - \frac{1}{\sqrt{2\pi}} \frac{e^{-\frac{\ln^2 z}{2}}}{\ln z}, \quad (2.6.32)$$

$$g(z) = -\ln(-\ln \mu(z)) \approx \frac{\ln^2 z}{2} + \ln \ln z + \ln \sqrt{2\pi}. \quad (2.6.33)$$

A kétszeres negatív logaritmussal definiált $g(z)$ függvény segítségével (2.4.3) és (2.4.4) alapján meghatározzuk a sztenderdizáció paramétereit, miszerint

$$b_N = g^{-1}(\ln N) \approx e^{\sqrt{2 \ln N}}, \quad (2.6.34)$$

$$a_N = \frac{1}{g'(b_N)} \approx \frac{b_N}{\ln b_N}, \quad (2.6.35)$$

majd abból (2.4.8) szerint a γ paramétert, majd az FSC amplitúdóját és γ' -t

$$\gamma_N = \frac{da_N}{db_N} \approx \frac{1}{\sqrt{2 \ln N}} \rightarrow \gamma = 0, \quad (2.6.36)$$

$$\epsilon_N = \gamma_N = \frac{1}{\sqrt{2 \ln N}}, \quad (2.6.37)$$

$$\gamma' = 0. \quad (2.6.38)$$

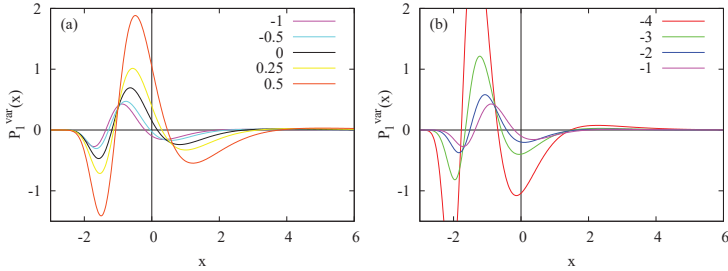
A konvergencia logaritmikus N -nel, de még a Gauss-típusú őseloszlásnál is lassabb, $1/\sqrt{\ln N}$ -nel arányos, az FSC függvénye ($\gamma = \gamma' = 0$) $P_1(x; 0, 0)$, az elsőrendű véges méret korrekció tehát

$$\text{FSC}_{\log \text{norm}} : \quad \epsilon_N P_1(x) = \frac{1}{\sqrt{2 \ln N}} P_1(x; 0, 0). \quad (2.6.39)$$

Végül említsük meg az FTG-eloszlást, mint őseloszlást. Ebből az őseloszlásból kiindulva nincs véges méret korrekció, az extrém eloszlás tetszőleges N -re a sztenderdizációt teljesítő átskálázás után a határeloszlással megegyezik. Ez ekvivalens azzal az állítással, miszerint az FTG-eloszlás az RG transzformáció fixpontja (l. (2.2.5) formula), a transzformáció hatása egy függvényre

pedig nem más, mint a határeloszláshoz való konvergencia. A fixpontot a transzformáció helybenhagyja, a sajátirányok mentén nem mozdtja el, csupán átskálázza, tehát nincs véges méret viselkedés.

Összefoglalva azt mondhatjuk, hogy a határeloszláshoz való konvergencia tipikusan lassú, N mintamérettel logaritmikus az FTG osztályon belül. Általánosított exponenciális őseloszlást vizsgálva ($\mu(z) \approx 1 - Be^{-Az^\delta}/z^\theta$) láttuk, hogy a véges méret korrekció amplitúdója $1/\ln N$ -nel arányos ha $\delta \neq 1$, és $1/\ln^2 N$ -tel ha $\delta = 1$. Elmondhatjuk továbbá, hogy az őseloszlásban szereplő hatványkitevő, θ vezető rendben nem befolyásolja a FSC amplitúdóját ha $\delta \neq 1$. Kivételt képeznek a lassú konvergencia alól az FTG-hez kellően közeli őseloszlások ($\gamma' < 0$), ilyen például a tiszta exponenciális őseloszlás ($\gamma' = -1$), amelyre az FSC amplitúdója $1/N$ -nel arányos.

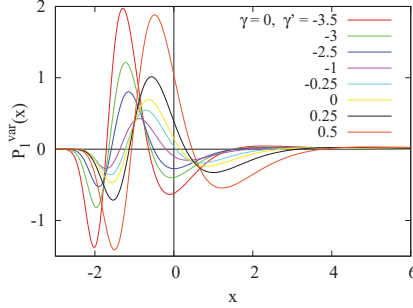


2.6.3. ábra. A sűrűségfüggvény első rendű véges méret korrekciójának függvénye $P_1^{\text{var}}(x; 0, \gamma') = \partial_x M_1^{\text{var}}(x; 0, \gamma')$ az FTG osztályban ($\gamma = 0$) különböző γ' értékekre a (2) sztenderdizációban, az ábrán $\gamma' = -4$ -től 0.5-ig. Nagy negatív γ' értékek felé haladva a függvény balra tolódik, amplitúdója $\gamma' = -1$ -ig csökken (a), utána ez a tendencia megfordul (b).

Általános esetben, tetszőleges γ' – vagyis tetszőleges konvergencia sebesség – mellett az FTG osztályban ($\gamma = 0$) az első rendű véges méret korrekció függvényének alakját (2.3.8) sajátfüggvény határozza meg, az integrált eloszlás korrekciója ezek alapján

$$M_1(x; 0, \gamma') = e^{-x-e^{-x}} \frac{1}{\gamma'^2} \left(1 + \gamma'x - e^{\gamma'x} \right) \quad (2.6.40)$$

az (1) sztenderdizációban. Áttérni a (2) típusú sztenderdizációra (2.5.8) sze-



2.6.4. ábra. A sűrűségfüggvény véges méret korrekciója FTG osztályra ($\gamma = 0$). Az előző ábrán tapasztalt tendencia összefoglalása. Itt is láthatjuk, hogy γ' abszolút értékét növelve a függvény balra tolódik, amplitúdója viszont nem monoton változik γ' -vel, $\gamma' = -1$ -nél minimuma van.

rint tudunk, ahol a transzformáció paraméterei

$$\begin{aligned} a &= \frac{\pi}{\sqrt{6}}, & a_1 &= -\frac{1}{\gamma'^2} \left[\gamma' a + \frac{1}{a} \Gamma(1 - \gamma') \left(\Psi(1 - \gamma') + b \right) \right], \\ b &= \gamma_E, & b_1 &= \frac{1}{\gamma'^2} \left(\Gamma(1 - \gamma') - 1 - \gamma' b \right), \end{aligned} \quad (2.6.41)$$

és $\Psi(x) = \Gamma'(x)/\Gamma(x)$. A variancia szerint sztenderdizált véges méret korrekció alakja tehát

$$\begin{aligned} M_1^{\text{var}}(x; 0, \gamma') &= \\ &= e^{-ax-b-e^{-ax-b}} \frac{1}{\gamma'^2} \left[\Gamma(1 - \gamma') \left(1 - \frac{1}{a} \left(\Psi(1 - \gamma') + b \right) x \right) - e^{\gamma'(ax+b)} \right]. \end{aligned} \quad (2.6.42)$$

A sűrűségfüggvényhez tartozó véges méret korrekciót az FTG osztályban, a 2.6.3. és a 2.6.4. ábra illusztrálja különböző γ' értékekre. A véges méret korrekciót γ' függvényében megvizsgálva azt láthatjuk, hogy az egyre növekvő negatív γ' értékek felé haladva a korrekció függvénye megtartva jellegzetes – négy szélsőértékkel rendelkező – alakját balra tolódik. Amplitúdója ezzel szemben nem viselkedik monoton módon, $\gamma' = -1$ -nél minimumot mutat. Ezt a tendenciát figyelhetjük meg a 2.6.3. ábra két oldalán illetve összegezve egy grafikonon a 2.6.4. ábrán.

A konvergencia – az RG szemléletben stabilitás – feltétele $\gamma' < 0$, ez γ' exponens definíciójából adódik (2.3.11). Azonban létezhetnek őseloszlások, melyek pozitív γ' értéket határoznak meg (pozitív γ' esetén érvényes korrekció függvényre példa a 2.6.3. ábra bal oldalán a $\gamma' = 0.5$ és 0.25 görbe). Az instabil irányok részletes vizsgálata megmutatta, hogy egy instabil sajátirány mentén elhelyezkedő őseloszlásból kiindulva az RG transzformáció hatására az eloszlás előbb eltávolodik majd visszatér a fix vonalhoz [27]. A $\gamma' = 0$ érték marginális stabilitást jelent, ami az általánosított exponenciális őseloszlás példáján demonstrált logaritmikus konvergenciának felel meg.

3. fejezet

SORREND STATISZTIKA

Egy változó extrém viselkedésének leírása során a legnagyobb érték eloszlásán túl más statisztikai jellemzők is érdekesek lehetnek. Ezt mutatja az extrémum körüli jellemzőkkel kapcsolatos korábbi tanulmányok sokfélesége is, úgymint az extrémum körüli állapotsűrűség [35, 36], átmeneti és visszatérési idők [37, 38] és rekord statisztikák [39]-[41] vizsgálata.

Ebben a fejezetben a $z_1, z_2, \dots, z_k, \dots$ sorozatot, vagyis az első, második, stb., k -adik legnagyobb értéket, és a köztük lévő távolságokat, vagyis a sorrend statisztikát vizsgáljuk. A sorrend statisztika elméletének legjobban ismert területe az iid változók esete [38, 42], ahol a maximum határeloszlásai kapcsán megismert klasszifikálás (FTG, FTF, FTW) megmarad, a k -adik maximum határeloszlása szintén három eloszlás-osztályt eredményez. A matematikai irodalom vizsgálta a gyengén korrelált rendszerek esetét is [38], erősen korrelált változókra azonban csak néhány partikuláris eredmény ismert [43].

Korrelált változók alaposabb megértése érdekében $1/f^\alpha$ zajok sorrend statisztikáját vizsgáljuk szimulációk és elméleti megfontolások segítségével. A korreláció mértéke az $1/f^\alpha$ zajoknál az α paraméterrel állítható, mivel $\alpha = 0$ az iid esetnek felel meg, míg $0 < \alpha < 1$ gyengén korrelált, $1 \leq \alpha \leq \infty$ tartomány pedig erősen korrelált jelet határoz meg. A k -adik távolság skálázásának szempontjából három különböző tartományt találunk a korreláció mértékének függvényében, melyeket az $\alpha = 1$ és $\alpha = 5$ határok választanak el egymástól.

Részletesebben vizsgáljuk az első és második legnagyobb érték közötti várható távolságot iid változókra, általános γ mellett, a három határeloszlás osztályt egységesen tárgyalva. Ez az érték az univerzális eloszlás osztályra jellemző, és ezért az empirikus adatok vizsgálata során alkalmas az osztály azonosítására. Az általános esetben felírt $N \rightarrow \infty$ határértéken túl meg-

vizsgáljuk az első-második közti távolság véges méret korrekcióját az FTG osztályban, különböző őseloszlások mellett, és az elméletet tovább bővítjük a következő fejezet végén véletlen mintaméret bevezetésével.

Különösen azért érdekes az első és második legnagyobb érték várható távolsága számunkra, mert az 5. fejezetben tárgyalt galaxis fényességek kapcsán ezt a mennyiséget hosszan elhúzódó vita övezi [65, 66, 68]. Klaszterekben vizsgálták a gap értékét, és az elméleti várakozásnál nagyobbak találták, ami arra engedett következtetni, hogy a klaszterek legnagyobb galaxisa nem felel meg a statisztikai maximumnak. Nagyobb adatbázisokat vizsgálva és körültekinthető extrém statisztikát alkalmazva a legújabb eredmények ezt cáfolni látszanak [69]. Kiszámoljuk a gap véges méret korrekcióját, mivel ez egyelőre tisztázatlan maradt, és később az 5. fejezet végén megvizsgáljuk, hogy az empirikus és az elméleti értékek eltérése származhat-e a véges méretből.

3.1. $1/f^\alpha$ zajok sorrend statisztikája

Periodikus ($z(t+T) = z(t)$), gaussi, $1/f^\alpha$ teljesítményspektrumú jeleket vizsgáltunk, amik α egyes értékei mellett jól ismert fizikai folyamatoknak felelnek meg, úgymint fehér zaj ($\alpha = 0$), $1/f$ zaj ($\alpha = 1$), véletlen bolyongás vagy Edwards-Wilkinson felület ($\alpha = 2$), véletlen gyorsulás vagy Mullins-Herring felület ($\alpha = 4$). A vizsgált mennyiség a zaj átlagos értéktől mért legmagasabb értéke (MRH) $z_1 = \max_t z(t) - \overline{z(t)}$ volt, aminek statisztikája utóbbi időben egyre nagyobb figyelmet kap [30, 32, 44], mi ezeket a kutatásokat általánosítjuk a sorrend statisztikára.

Egy $z(t)$ -vel jelölt T periódusú $1/f^\alpha$ jel konfigurációjának sulya

$$\mathcal{P}[z(t)] \propto e^{-S[z(t)]}, \quad (3.1.1)$$

ahol a $z(t)$ jel $S[z(t)]$ hatásfüggvényét

$$S[\{c_n\}; \alpha] = (2\pi)^\alpha T^{1-\alpha} \sum_{n=1}^{N/2} n^\alpha |c_n|^2 \quad (3.1.2)$$

a diszkrét Fourier-kifejtés

$$z(t) = \sum_{n=-N/2+1}^{N/2} c_n e^{2\pi i n t/T}, \quad c_n^* = c_{-n}, \quad (3.1.3)$$

c_n együtthatói határozzák meg, ahol N páros, pozitív egész, és c_n fázisa egyenletes eloszlású véletlen változó a $[0, 2\pi]$ intervallumon [32, 45]. A formulából jól látszik, hogy a Fourier térben a c_n együtthatók független, de (az $\alpha = 0$ eset kivételével) különböző eloszlású változók. Gauss-eloszlásból

származnak, nulla várható értékkel, de szórásuk különbözik, $\langle |c_n|^2 \rangle \propto 1/n^\alpha$. A valós térbeli jel, $z(t)$, különböző időpontokbeli értékei viszont korreláltak lesznek (az $\alpha = 0$ eset kivételével), és a korreláció mértéke α -val nő. Az $0 < \alpha < 1$ tartományban a jel magasságértékei között hatványszerűen lecsengő, de véges integrállal rendelkező korreláció van, míg $1 \leq \alpha \leq \infty$ tartományban a korreláció integrálja divergál, vagyis a fluktuáció is divergálnak. Ezt a két tartományt a továbbiakban a gyenge illetve az erős korreláció tartományának nevezzük.

A Fourier-együtthatók fenti tulajdonsága miatt az $1/f^\alpha$ jelek szimulálása egyszerű. A megfelelő eloszlással generáljuk a c_n amplitúdókat és fázisukat, majd Fourier transzformáljuk a jelet. Ezzel a módszerrel generálva a $z(t)$ időbeli jel is diszkrét lesz, mivel véges számú (N) Fourier-módusból állítottuk elő, vagyis folytonos jel helyett N db $z_n = z(nT/N)$, $n = 0, 1, \dots, N-1$ korrelált változót kapunk. Nagyszámú jelet generáltunk (10^6), és a rendezett minták alapján meghatároztuk a k -adik legnagyobb érték eloszlását, várható értékét. Mindezt különböző α értékekre a korreláció, és különböző N értékre az extrém érték elmélet szerinti $N \rightarrow \infty$ konvergencia vizsgálatára. Megvizsgáltuk a véges méretű mintában meghatározott $\langle z_k \rangle_N$ átlagok és $\langle z_k \rangle_N - \langle z_{k+1} \rangle_N$ távolságok N -től való függését, és azt kiskálázva, az $N \rightarrow \infty$ limeszben nem divergáló távolságokkal foglalkoztunk. A $k = 1, 2, 5$ és 10-edik legnagyobb érték szimulált valószínűségi sűrűségfüggvényét mutatja a 3.1.1. ábra, változó korreláció, $\alpha = 0, 0.5, 1$ és 2 mellett, az $x = (z - \langle z_1 \rangle)/\sigma_1$ átskálázott változóban, ahol $\langle z_1 \rangle$ a legnagyobb érték átlaga és σ_1 annak szórása.

Az alábbiakban részletesen tárgyaljuk a különböző α tartományokat. Előtte röviden összefoglaljuk, hogy milyen mennyiségeket vizsgálunk és mit mutatnak a szimulációk és az egyéb elméleti megfontolások. Az átlagos távolság a k -adik és a $k+1$ -edik legnagyobb érték között, $d_k = \langle x_k - x_{k+1} \rangle$ (ahol d_k az N -függést már nem tartalmazó távolság), a következőképpen skálázik k -val

$$d_k \sim \begin{cases} k^{-1} & 0 \leq \alpha < 1, \\ k^{(\alpha-3)/2} & 1 < \alpha < 5, \\ k & 5 < \alpha \leq \infty. \end{cases} \quad (3.1.4)$$

Vagyis három különböző skálázási tartományt különböztetünk meg, amelyek határán folytonosan illeszkedik az exponens. A távolságok skálázása arra enged következtetni, hogy a k -adik legnagyobb várható értéke

$$\varepsilon_k = \langle x_1 - x_k \rangle \sim \begin{cases} \ln k & 0 \leq \alpha < 1, \\ k^{(\alpha-1)/2} & 1 < \alpha < 5, \\ k^2 & 5 < \alpha \leq \infty. \end{cases} \quad (3.1.5)$$

Ez az eredmény $1 < \alpha \leq \infty$ -re összhangban van az extrémum közelében számolt állapotsűrűségből kapható exponensekkel [36].

3.1.1. Független változók, $\alpha = 0$

Az $1/f^\alpha$ zajt meghatározó valószínűségből látszik, hogy ha $\alpha = 0$, akkor tulajdonképpen N db iid Gauss-eloszlású változó sorrend statisztikáját vizsgáljuk. A k -adik legnagyobb érték várható értéke és szórása ekkor N -nel a következőképpen viselkedik

$$\langle z_k \rangle \sim \sqrt{\ln N} \quad , \quad \sigma_k \sim 1/\sqrt{\ln N} \quad , \quad (3.1.6)$$

és az $x = (z - \langle z_1 \rangle)/\sigma_1$ átskálázott változóban a k -adik maximum határeloszlása $N \rightarrow \infty$ esetén [38]

$$P_{(k)}(x) = \frac{1}{(k-1)!} \exp(-k\tilde{x} - e^{-\tilde{x}}) \quad , \quad \tilde{x} = ax + \gamma_E \quad , \quad (3.1.7)$$

ahol $a = \pi/\sqrt{6}$ és γ_E az Euler konstans. Az elméleti határeloszlásokat és a véges mérettel szimulált eloszlásokat veti össze a 3.1.1.(a) ábra. A logaritmikusan lassú, $\ln N$ -el arányos konvergencia miatt a véges méret korrekció még $N = 16384$ esetén is jelentős, és k növekvő értékeire a korrekció mértéke láthatóan tovább nő.

(3.1.7) eloszlás alapján meghatározhatjuk a k -adik távolság elméleti határértékét

$$d_k = \langle x_k \rangle - \langle x_{k+1} \rangle = \frac{1}{ak} \quad , \quad (3.1.8)$$

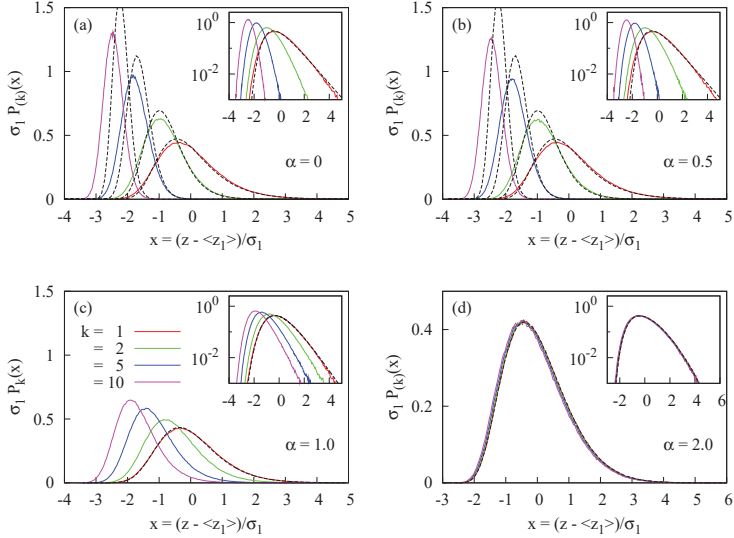
és egyszerű megfontolás alapján nagy k esetén az $\langle x_k \rangle$ legnagyobb értékek spektrumát

$$\varepsilon_k \equiv \langle x_1 \rangle - \langle x_k \rangle = \sum_{\ell=1}^{k-1} d_\ell = \frac{1}{a} \sum_{\ell=1}^{k-1} \frac{1}{\ell} \approx \frac{1}{a} \ln k \quad . \quad (3.1.9)$$

Ezeket az elméleti határeloszlásból kapott eredményeket ugyanakkor a szimulációkban tapasztalt skálázás alapján is le tudjuk vezetni néhány egyszerű megfontolás segítségével. A rendezett mintában $k = 1$ -től $N/2$ -ig összegezve a távolságokat, az őseloszlás mediánja és a legnagyobb érték közötti távolságot kapjuk

$$\sum_{k=1}^{N/2} \langle z_k - z_{k+1} \rangle \approx \langle z_1 \rangle - \langle z_{N/2} \rangle \quad . \quad (3.1.10)$$

A szimulációkból azt tapasztaljuk, hogy az N mintamérettel $\langle z_1 \rangle \sim \sqrt{\ln N}$



3.1.1. ábra. A k -adik legnagyobb érték valószínűségi sűrűségfüggvénye $P_k(x)$, $k = 1, 2, 5, 10$ -re (jobbról balra), szimulációk alapján ($N = 16384$), átskalázva $x = (z - \langle z_1 \rangle) / \sigma_1$ szerint (a jobb felső sarkokban ugyanez félogaritmikuskálán). Az $\alpha = 0$, iid esetben tetszőleges k -ra ismerjük az egzakt határeloszlást (fekete szaggatott görbék), amik minden jel szerint kiterjeszthetők az $0 < \alpha < 1$ gyengén korrelált tartományra, láthatjuk, hogy a véges méret korrekció jelentős $\alpha < 1$ -re. Az $\alpha = 1$ és 2 erősen korrelált esetben csak a $k = 1$ határeloszlás ismert egzaktul, viszont $\alpha \geq 1$ -re a konvergencia gyors, s az $N = 16384$ mintamérettel szimulált eloszlás gyakorlatilag megegyezik a határeloszlással.

és $\langle z_k - z_{k+1} \rangle \sim d_k / \sqrt{\ln N}$ szerint divergál, ezeket az összefüggéseket felhasználva

$$\sum_{k=1}^{N/2} d_k / \sqrt{\ln N} \approx \sqrt{\ln N}, \quad \text{vagyis} \quad \sum_{k=1}^{N/2} d_k \approx \ln N, \quad (3.1.11)$$

mivel a medián nem divergál N -el, $\langle z_{N/2} \rangle \sim \mathcal{O}(1)$. Egyszerű hatványfüggést, $d_k \sim k^\delta$, feltételezve (3.1.11) alapján megállapíthatjuk az elméleti eredménnyel egybeeső skálázást

$$d_k \sim k^{-1}. \quad (3.1.12)$$

3.1.2. Korrelált változók, $\alpha \neq 0$

Gyengén korrelált változók, $0 < \alpha < 1$

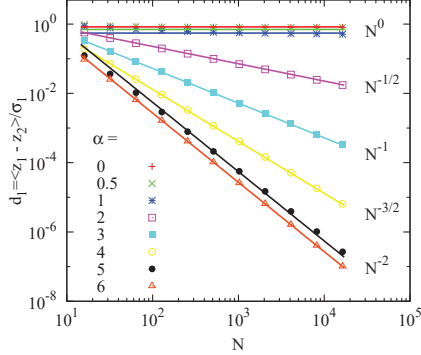
Gyenge korreláció jelenlétében, ami az $1/f^\alpha$ zajok esetén az $0 < \alpha < 1$ tartománynak felel meg, az iid határeloszlások nem változnak. $k = 1$ esetén ismerjük az egzakt eredményt, a legnagyobb érték eloszlása az FTG-hez tart [28], ennek a tudatában és a szimulációkban tapasztalt konvergencia alapján azt várjuk, hogy az iid változóknál megismert $P_{(k)}(x)$ határeloszlások érvényesek maradnak tetszőleges k -ra. A véges méret korrekció azonban továbbra is jelentős, logaritmikus vagy annál lassabb [32]. A 3.1.1.(b) ábrán látjuk, $\alpha = 0.5$ -re, hogy a szimulált eloszlások az $\alpha = 0$ esethez nagyon hasonlóak. A szimulációk és az elméleti eredmények alapján tudjuk, hogy $\langle z_1 \rangle \sim \sqrt{\ln N}$ és $\sigma_1 \sim 1/\sqrt{\ln N}$ -el az iid esettel megegyezően skálázik, és ezért mivel a szórással átskálázott távolságok egységnyi nagyságrendben vannak (l. 3.1.1.(b) ábra), $\langle z_k - z_{k+1} \rangle \sim 1/\sqrt{\ln N}$. Ezt mutatja a 3.1.2. ábra is, ahol a skálázott első távolságot, $\langle z_1 - z_2 \rangle / \sigma_1$, ábrázoljuk, ami valóban $\mathcal{O}(1)$. A szimulált távolságokat N függvényében ábrázolva is az $\langle z_k - z_{k+1} \rangle \sim 1/\sqrt{\ln N}$ skálázást tapasztaljuk. Az $\alpha = 0$ esetben végzett (3.1.11) megfontolás ezért itt is ugyanarra a skálázásra vezet, miszerint

$$d_k \sim k^{-1}. \quad (3.1.13)$$

Amennyiben a távolságokat k függvényében ábrázoljuk, a log-log skálán illesztett egyenes iránytangense alátámasztja ezt a skálázást (3.1.3. ábra).

Nemtriviális skálázás tartománya, $1 \leq \alpha < 5$

Amennyiben $\alpha \geq 1$, az $1/f^\alpha$ zaj $z(t)$ magasság értékei erősen korrelált változók lesznek, és a fluktuációik divergálnak. A sorrend statisztika jellemzése



3.1.2. ábra. Az első és második legnagyobb érték közötti skálázott távolság, $\langle z_1 - z_2 \rangle / \sigma_1$, függése az N mintamérettől. Az α -val parametrizált korreláció változtatásával három skálázási tartományt különböztethetünk meg.

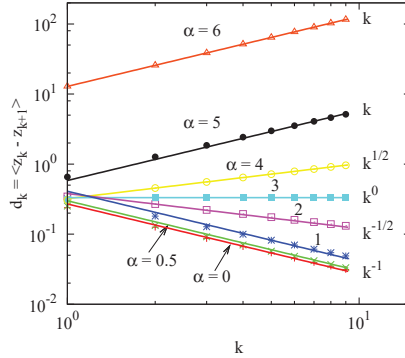
szempontjából az $1 \leq \alpha \leq \infty$ tartományt két részre bontja az $\alpha = 5$ határ, mivel $\alpha \geq 5$ -re a zaj második deriváltja folytonossá válik. Előbb tekintsük az $1 \leq \alpha < 5$ tartományt.

Az $\alpha = 1$ eset, mivel két különböző skálázási tartományt választ el, nehezen tisztázható a lehetséges logaritmikus korrekciók miatt. A szimulációkból meghatározott $P_{(k)}(x)$ eloszlások (3.1.1.(c) ábra) az $\alpha < 1$ esetekhez képest láthatóan közelebb kerültek egymáshoz, mind a skála mind az eltolás szempontjából. A legnagyobb érték átlaga $\langle z_1 \rangle \sim \ln N$ -nel skálázik, és ismerjük az elméleti határeloszlást [44], aminek alakja az átlaggal eltoltt $y = z_1 - \langle z_1 \rangle$ változóban

$$P(y) = [2e^{-y/2} K_1(2e^{-y/2})]', \quad (3.1.14)$$

ahol K_1 a módosított Bessel-függvény. A legnagyobb értékre szimulált, véges N eloszlások és a határeloszlás jó egyezése azt mutatja, hogy a konvergencia $\alpha = 1$ esetén, az erősen korrelált változókra jellemző módon, gyors. Mivel $\sigma_1 \sim \mathcal{O}(1)$, a 3.1.1.(c) ábráról leolvashatjuk, hogy a távolságok, $d_k \sim \mathcal{O}(1)$ egységnyiek.

A skálázás megváltozása $\alpha = 1$ -nél szembeűnő, mivel $\alpha > 1$ -re a $P_{(k)}(x)$ eloszlások a 3.1.1.(d) ábra skáláján egybeesnek, vagyis a távolságok – az $\alpha < 1$ esetektől eltérően – σ_1 -nél kisebb nagyságrendűek lesznek. A min-



3.1.3. ábra. A k -adik távolság átlagának, $d_k = \langle z_k - z_{k+1} \rangle$ -nak a skálázása k -val szimulációk alapján $N = 16384$ -re, $\alpha = 0, 0.5, 1, \dots, 6$ -ra. $d_k \sim k^\delta$ összefüggést feltételezve meghatároztuk az exponenseket az egyes α értékek mellett, az $\alpha = 1$ és $\alpha = 5$ határok kivételével (ahol logaritmikus korrekciók jelenhetnek meg) az egyenesek rendkívül jól illeszkednek a szimulációk eredményeire.

taméréttel való konvergencia drasztikusan felgyorsul, hatványszerű lesz [32], a gyors konvergenciát jól illusztrálja a 3.1.1.(d) ábrán az $\alpha = 2$ esete. Az elméleti határeloszlással, ami $\alpha = 2$ -re az Airy-függvény [30, 34], már kis N értékre is egybeesik a véges mérettel simulált extrém eloszlás. A szimulációk szerint a skálázott távolság $d_k/\sigma_1 \sim N^{-(\alpha-1)/2}$ -vel függ a mintamérettől (l. a 3.1.2. ábra), ugyanakkor ismert, hogy $\sigma_1 \sim N^{(\alpha-1)/2}$ [32], emiatt $\langle z_k - z_{k+1} \rangle \sim \mathcal{O}(1)$, vagyis itt nincs szükség egy N -től függő átskálázásra, hogy véges határértéket kapjunk. A távolságokat k függvényében ábrázolva hatványszerű összefüggést láthatunk (3.1.3. ábra), ami az illesztett exponensekkel összhangban a

$$d_k \sim k^{(\alpha-3)/2} \quad (3.1.15)$$

skálázásra enged következtetni. Valóban, a korábbi elméleti meggondolást megismételve, $\langle z_1 \rangle \sim N^{(\alpha-1)/2}$ ismeretében

$$\sum_{k=1}^{N/2} d_k \approx \langle z_1 \rangle \sim N^{(\alpha-1)/2}, \quad (3.1.16)$$

és $d_k \sim k^\delta$ hatvánfüggést feltételezve a fenti (3.1.15) skálázást kapjuk. Ez-

után (3.1.15) integrálásával nagy k mellett megkapjuk a rendezett minta várható értékeinek spektrumát

$$\epsilon_k \sim k^{(\alpha-1)/2}. \quad (3.1.17)$$

Nagy α limit, $5 \leq \alpha \leq \infty$

Az $5 \leq \alpha \leq \infty$ tartományt – az $1 \leq \alpha < 5$ -höz hasonlóan – az erős korreláció, határeloszlásokhoz való gyors konvergencia és a $\langle z_1 \rangle \sim \sigma_1 \sim N^{(\alpha-1)/2}$ N -függés jellemzi. A különbség a sorrend statisztika szempontjából, hogy $\alpha \geq 5$ -re a jel második deriváltja folytonossá válik. Mivel a nagy α limitben csak egy módus dominálja a jelet, a legnagyobb értékek a maximum, $z(t_1)$, körül fognak elhelyezkedni, és ezért a sorrend statisztika meghatározható a

$$z(t_1 + \Delta) = z_1 + \frac{1}{2}z''(t_1)\Delta^2 + O(\Delta^3) \quad (3.1.18)$$

sorfejtés segítségével, ahol $\Delta = \tau, 2\tau, \dots, (k-1)\tau$ a második, harmadik, stb. legnagyobb értéknek megfelelően. Mivel a legnagyobb érték $\langle z_1 \rangle \sim N^{(\alpha-1)/2}$ módon viselkedik, z'' tipikus értékének becslése szerint

$$[|(z'')^2|]^{1/2} \sim \langle |z''(t_1)| \rangle \sim N^{(\alpha-5)/2} \sim \langle z_1 \rangle N^{-2}. \quad (3.1.19)$$

A d_k távolság (3.1.18) sorfejtéssel

$$\langle z(t_1 + k\tau) - z(t_1 + (k+1)\tau) \rangle \approx \langle |z''(t_1)| \rangle \tau^2 k \sim N^{(\alpha-5)/2} k, \quad (3.1.20)$$

miel $\tau = T/N \sim \mathcal{O}(1)$. Ezt az N -függést igazolják a szimulációk is, ahol $d_k/\sigma_1 \sim N^{-2}$ (l. 3.1.2. ábra), mivel tudjuk hogy $\sigma_1 \sim N^{(\alpha-1)/2}$. Az N függést eltávolítva a távolságokból a sorfejtés eredménye azt mutatja, hogy

$$d_k \sim k, \quad (3.1.21)$$

ami egybecseng a szimulációk szerint az $5 \leq \alpha \leq \infty$ tartományban tapasztalt skálázással (3.1.3. ábra), és ebből következően felírhatjuk a spektrumot

$$\epsilon_k \sim k^2. \quad (3.1.22)$$

3.2. Az első és második legnagyobb érték várható távolsága iid változókra

Az első és a második legnagyobb érték várható távolságát, $d_1 = \langle x_1 \rangle - \langle x_2 \rangle$ -ot vizsgáljuk részletesebben ebben az alfejezetben, független (iid) változókra. Meghatározzuk $N \rightarrow \infty$ határértékét a három határeloszlás osztályra egyélesen, és vezető rendű véges méret korrekcióját az FTG osztályra.

3.2.1. A távolság határértéke általános esetben

Amint azt a 2. fejezet elején láttuk, az első legnagyobb érték, a maximum eloszlása, ha z -t egy lineáris transzformációval átskalázzuk ($z = a_N x + b_N$) az (1) sztenderdizációnak megfelelően

$$M_{(1)}(x, N) = \mu(a_N x + b_N)^N \rightarrow M_{(1)}(x) = e^{-(1+\gamma x)^{-1/\gamma}}. \quad (3.2.1)$$

A k -adik legnagyobb érték eloszlását általánosan meghatározhatjuk egy egyszerű megfontolás segítségével. Keressük annak a valószínűségét, hogy a $\rho(z)$ öloszlásból húzott N darab érték közül a k -adik legnagyobb az $(z, z + dz)$ intervallumba esik. Vagyis az N értékből $k - 1$ darab lesz z -nél nagyobb, ennek a valószínűsége $[1 - \mu(z)]^{k-1}$, $N - k$ darab lesz z -nél kisebb, $\mu(z)^{N-k}$ valószínűséggel, és egy lesz a $(z, z + dz)$ intervallumban, $\rho(z)dz$ valószínűséggel, mindezt szorozzuk a megvalósulás kombinatorikus súlyával, vagyis

$$P_{(k)}^{\max}(z, N)dz = N \binom{N-1}{k-1} \rho(z)dz \left(1 - \mu(z)\right)^{k-1} \mu(z)^{N-k}. \quad (3.2.2)$$

A második legnagyobb érték valószínűségi sűrűségét $k = 2$ helyettesítéssel kapjuk

$$P_{(2)}^{\max}(z, N) = N(N-1)\rho(z) \left(1 - \mu(z)\right) \mu(z)^{N-2}, \quad (3.2.3)$$

ennek integrálásával pedig az integrált eloszlásfüggvényét

$$M_{(2)}^{\max}(z, N) = N\mu(z)^{N-1} - (N-1)\mu(z)^N = \mu(z)^N \left[1 + N \left(\frac{1}{\mu(z)} - 1\right)\right], \quad (3.2.4)$$

ahol végrehajtva a szokásos átskalázást ($z = a_N x + b_N$), az $N \rightarrow \infty$ limeszben

$$N \left(\frac{1 - \mu(z)}{\mu(z)}\right) \rightarrow (1 + \gamma x)^{-1/\gamma} \quad (3.2.5)$$

felhasználásával kapjuk, hogy a második legnagyobb érték integrált eloszlása

$$M_{(2)}(x, N) \rightarrow M_{(2)}(x) = e^{-(1+\gamma x)^{-1/\gamma}} \left[1 + (1 + \gamma x)^{-1/\gamma}\right]. \quad (3.2.6)$$

(3.2.1) illetve (3.2.6) deriválásával az első és második legnagyobb érték sűrűségeloszlása az $N \rightarrow \infty$ limeszben

$$P_{(1)}(x) = (1 + \gamma x)^{-1/\gamma-1} e^{-(1+\gamma x)^{-1/\gamma}}, \quad (3.2.7)$$

$$P_{(2)}(x) = (1 + \gamma x)^{-2/\gamma-1} e^{-(1+\gamma x)^{-1/\gamma}}, \quad (3.2.8)$$

ahol $1 + \gamma x \geq 0$. Az FTG osztályra ($\gamma = 0$)

$$P_{(1)}(x) = e^{-x-e^{-x}} \quad P_{(2)}(x) = e^{-2x-e^{-x}}. \quad (3.2.9)$$

Látható, hogy az első legnagyobb érték eloszlása segítségével a következőképpen írható a második eloszlása

$$M_{(2)}(x) = M_{(1)}(x) + (1 + \gamma x)P_{(1)}(x), \quad (3.2.10)$$

ezt felhasználva a második maximum várható értéke

$$\begin{aligned} \langle x_2 \rangle &= \int_{x_{inf}}^{x_{sup}} x P_{(2)}(x) dx = \int_{x_{inf}}^{x_{sup}} x \left[(1 + \gamma)P_{(1)}(x) + (1 + \gamma x)P'_{(1)}(x) \right] dx = \\ &= (1 + \gamma) \langle x_1 \rangle + \left[x(1 + \gamma x)P_{(1)}(x) \right]_{x_{sup}}^{x_{inf}} - \int_{x_{sup}}^{x_{inf}} (1 + 2\gamma x)P_{(1)}(x) dx, \\ \langle x_2 \rangle &= (1 - \gamma) \langle x_1 \rangle - 1. \end{aligned} \quad (3.2.11)$$

Vagyis az első és a második legnagyobb érték várható távolsága az (1) sztenderdizációban

$$d_1 = \langle x_1 \rangle - \langle x_2 \rangle = \gamma \langle x_1 \rangle + 1, \quad (3.2.12)$$

ami az FTG osztályra $d_1 = 1$. A (2) sztenderdizációra, ami a gyakorlat szempontjából érdekes, az $\langle x_1 \rangle$ átlagot levonva és a szórással leosztva tudunk áttérni, ami a távolság szempontjából egy $1/\sigma_1$ átskálázást jelent. Az FTG osztályra ez könnyen felírható

$$d_1^{\text{var}} = \frac{1}{\sigma_1} = \frac{\sqrt{6}}{\pi}. \quad (3.2.13)$$

Általános γ -t figyelembe véve az első legnagyobb várható értéke és szórása

$$\langle x_1 \rangle = \frac{\Gamma(1 - \gamma) - 1}{\gamma}, \quad (3.2.14)$$

$$\sigma_1 = \frac{\sqrt{\Gamma(1 - 2\gamma) - \Gamma^2(1 - \gamma)}}{|\gamma|}, \quad (3.2.15)$$

vagyis az első és a második maximum távolsága az (1) sztenderdizációban

$$d_1 = \Gamma(1 - \gamma), \quad (3.2.16)$$

és a (2) sztenderdizációban

$$d_1^{\text{var}} = \frac{|\gamma| \Gamma(1 - \gamma)}{\sqrt{\Gamma(1 - 2\gamma) - \Gamma^2(1 - \gamma)}}. \quad (3.2.17)$$

3.2.2. A távolság véges méret korrekciója az FTG osztályra

A gyakorlatban való alkalmazhatóság érdekében érdemes megvizsgálni az első és második legnagyobb érték távolságának véges méret korrekcióját is a leggyakrabban előforduló FTG osztály esetén. A második legnagyobb érték eloszlásának sűrűségfüggvénye véges N mintaméret mellett (3.2.3) szerint

$$P_{(2)}^{\max}(z, N) = N(N-1)\rho(z)\left(1 - \mu(z)\right)\mu(z)^{N-2}, \quad (3.2.18)$$

ezzel a második legnagyobb várható értéke

$$\langle z_2 \rangle = \int_{-\infty}^{\infty} z P_{(2)}^{\max}(z, N) dz, \quad (3.2.19)$$

amit vezető rendig sorbafejtünk nagy N értékekre a véges méret korrekció meghatározása érdekében. Mivel korábban láttuk, hogy az FTG vonzási tartományába tartozó őseloszlások két fő csoportra bonthatók a határeloszláshoz való konvergencia sebessége szempontjából, nevezetesen a logaritmikus ($\gamma' = 0$) és a hatványszerű (pl. $\gamma' = -1$) konvergencia csoportjára, ezért az első-második távolság véges méret korrekcióját is külön kell tárgyalnunk erre a két esetre.

Általánosított exponenciális őseloszlás esetén

Először tekintsünk egy, az FTG osztályba tartozó és az alkalmazásokban gyakran előforduló eloszlás családot, az általánosított exponenciális őseloszlás esetét. Az előző fejezetben láttuk, hogy az aszimptotikája szerint általában $\gamma = \gamma' = 0$, és a határeloszláshoz való konvergencia logaritmikus. Azt is tudjuk, hogy a határeloszláshoz tartozó FSC függvénye is univerzális a γ, γ' értékeivel meghatározott osztályon belül, csupán ϵ_N , a korrekció amplitúdója függ az őseloszlás pontosabb alakjától. Ezért várakozásaink szerint elég egy speciális esetben kiszámolni a távolság véges méret korrekcióját.

A $\gamma = \gamma' = 0$ eloszlás család demonstrálására a Rayleigh-eloszlással végezzük el a sorfejtést. A $\mu(z) = 1 - e^{-z^2}$ őseloszlást vesszük, és (3.2.19) kiszámolása során az integrálban a $z = \sqrt{y + \ln N}$ változócsere után sorbafejtünk, ebben az esetben $1/\ln N$ hatványai szerint. Az egyes tagok $P_{(2)}^{\max}(z, N)$

függvényében vezető rendig

$$\begin{aligned} N(N-1)\rho(z)\left(1-\mu(z)\right) &= N(N-1)2ze^{-2z^2} = 2\sqrt{y+\ln N}e^{-2y} + \mathcal{O}(1/N) \\ &\approx 2\sqrt{\ln N}\left(1 + \frac{y}{2\ln N} - \frac{y^2}{8\ln^2 N}\right)e^{-2y}, \end{aligned} \quad (3.2.20)$$

$$\mu(z)^{N-2} = \left(1 - e^{-z^2}\right)^{N-2} = \frac{e^{N\ln\left(1 - \frac{e^{-y}}{N}\right)}}{\left(1 - \frac{e^{-y}}{N}\right)^2} \approx e^{-e^{-y}} + \mathcal{O}(1/N), \quad (3.2.21)$$

ezáltal a második maximum várható értéke

$$\begin{aligned} \langle z_2 \rangle &= \int_{-\infty}^{\infty} P_{(2)}^{\max}(\sqrt{y+\ln N}, N) \frac{dy}{2} = \\ &= \sqrt{\ln N} \int_{-\infty}^{\infty} \left(1 + \frac{y}{2\ln N} - \frac{y^2}{8\ln^2 N}\right) e^{-2y-e^{-y}} dy. \end{aligned} \quad (3.2.22)$$

Itt a vezető renden túl a következő korrekciós tagra is szükség van, mivel $1/\sigma_1$ -el való átskálázás után az is járulékot fog adni

$$\langle z_2 \rangle = \sqrt{\ln N} + \frac{1}{2\sqrt{\ln N}}(\gamma_E - 1) - \frac{1}{8\ln^{3/2} N} \left(\gamma_E^2 - 2\gamma_E + \frac{\pi^2}{6} \right). \quad (3.2.23)$$

Az első maximum várható értéke analóg módon számolva egy korábbi eredményünk alapján [47]

$$\langle z_1 \rangle = \sqrt{\ln N} + \frac{1}{2\sqrt{\ln N}}\gamma_E - \frac{1}{8\ln^{3/2} N} \left(\gamma_E^2 + \frac{\pi^2}{6} \right), \quad (3.2.24)$$

a kettő távolsága tehát

$$\langle z_1 \rangle - \langle z_2 \rangle = \frac{1}{2\sqrt{\ln N}} - \frac{1}{4\ln^{3/2} N} \gamma_E. \quad (3.2.25)$$

Ez az eredmény a távolságot az ösloszlás eredeti skáláján adja meg, mivel azonban a Rayleigh-ösloszlás esetén a skála is változik N értékével, ezért az (1) sztenderdizációra $1/a_N$ -nel szorozva térhetünk át. A (2.6.4) általános formula azonban csak az a_N nulladrendű tagját adja meg, mivel jelen esetben $\theta = 0$, így a vezető korrekció meghatározásához (2.6.2) sorfejtésben egy taggal tovább kell mennünk

$$g(z) = -\ln(-\ln \mu(z)) \approx z^2 - \frac{e^{-z^2}}{2}, \quad (3.2.26)$$

vagyis az a_N ebből következően

$$a_N = \frac{1}{g'(b_N)} \approx \frac{1}{2b_N} \left(1 - \frac{e^{-b_N^2}}{2} \right), \quad (3.2.27)$$

ahol (2.6.3) alapján $b_N = \sqrt{\ln N}$, ezt behelyettesítve láthatjuk, hogy az

extrém eloszlás $1/2\sqrt{\ln N} + \mathcal{O}(1/N)$ -nel szűkül. Mivel az $\mathcal{O}(1/N)$ korrekció $1/\ln N$ mellett elhanyagolható, a távolság az (1) sztemderdizációnak megfelelően

$$d_1 = 1 - \frac{1}{2 \ln N} \gamma_E. \quad (3.2.28)$$

Az első maximum szórása [47] az eredeti skálán

$$\sigma_1 = \frac{1}{2\sqrt{\ln N}} \frac{\pi}{\sqrt{6}} \left[1 - \frac{1}{2 \ln N} \left(\frac{6}{\pi^2} \zeta(3) + \gamma_E \right) \right], \quad (3.2.29)$$

ahol $\zeta(3)$ az Apéry-állandó. A szórással osztva az eredeti skálán mért első-második távolságot, felírhatjuk a távolságot a (2) sztemderdizációban

$$d_1^{\text{var}} = \frac{\sqrt{6}}{\pi} \left[1 + \frac{1}{2 \ln N} \frac{6}{\pi^2} \zeta(3) \right]. \quad (3.2.30)$$

Mivel (2.6.7) alapján a Rayleigh-őseloszlásra $\epsilon_N = -1/2 \ln N$ a véges méret korrekciót jellemző amplitúdó, ezért ϵ_N bevezetésével általánosítjuk az eredményt a $\gamma = \gamma' = 0$ őseloszlások esetére. Az első és második legnagyobb érték távolsága az elsőrendű véges méret korrekciót is figyelembe véve, ha az őseloszlás aszimptotikája szerint $\gamma = \gamma' = 0$, az (1) és (2) sztemderdizáció szerint tehát

$$d_1 = 1 + \epsilon_N \gamma_E, \quad (3.2.31)$$

$$d_1^{\text{var}} = \frac{\sqrt{6}}{\pi} \left[1 - \epsilon_N \frac{6}{\pi^2} \zeta(3) \right], \quad (3.2.32)$$

ahol az ϵ_N amplitúdót (2.6.7) határozza meg az általánosított exponenciális őseloszlás esetén, általános esetben a 2.4. fejezet alapján számolható.

Hogy ϵ_N bevezetése mennyire jogos, azt a következő okfejtéssel láthatjuk be. Felhasználva (2.3.13)-et a legnagyobb érték ($k=1$) véges méretű extrém eloszlása a következő módon közelíthető

$$P_{(1);N}(x) = P_{(1)}(x; \gamma) + \epsilon_N P_{(1);1}(x, \gamma, \gamma'). \quad (3.2.33)$$

Ezért az első maximum várható értéke az elsőrendű véges méret korrekciót is figyelembe véve, általános γ' -re az FTG osztályban ($\gamma = 0$) az (1) sztemderdizációban

$$\langle x_1 \rangle = \int x P_{(1);N}(x) dx = \gamma_E + \epsilon_N \int x P_{(1);1}(x, 0, \gamma') dx. \quad (3.2.34)$$

Amennyiben feltesszük, hogy a $k=1$ -gyel analóg módon, a második maximum véges méretű eloszlása is felírható (3.2.33) mintájára a $P_{(2)}(x; \gamma)$ határeloszlás és egy véges méret korrekció összegeként, ami egy γ és γ' -től függő

függvény ϵ_N amplitúdóval¹

$$P_{(2);N}(x) = P_{(2)}(x; \gamma) + \epsilon_N P_{(2);1}(x, \gamma, \gamma'), \quad (3.2.35)$$

akkor a második maximum várható értéke

$$\langle x_2 \rangle = (\gamma_E - 1) + \epsilon_N \int x P_{(2);1}(x, 0, \gamma') dx. \quad (3.2.36)$$

Vagyis az első-második közötti távolság a (1) sztenderdizációban

$$d_1 = 1 + \epsilon_N f(\gamma'), \quad (3.2.37)$$

ahol

$$f(\gamma') = \int x \left(P_{(1);1}(x, 0, \gamma') - P_{(2);1}(x, 0, \gamma') \right) dx \quad (3.2.38)$$

csak a γ' paramétertől függ, vagyis az összes olyan őseloszlásra megegyezik, amelyre $\gamma = \gamma' = 0$, (3.2.37) és (3.2.31) összevetése alapján értéke $f(0) = \gamma_E$.

Tiszta exponenciális őseloszlás esetén

A tiszta exponenciális őseloszlás kivétel az FTG osztályban tapasztalt tipikus logaritmikus konvergencia alól, aszimptotikája alapján $\gamma = 0$ és $\gamma' = -1$, ezért véges méret viselkedése különbözik a fentebb számolt általánosított exponenciálistól, így a távolság korrekciója is attól eltérő lesz.

Tekintsük az exponenciális őseloszlást, vagyis $\mu(z) = 1 - e^{-z}$. A $z = y + \ln N$ változócsere után $P_{(2)}^{\max}(z, N)$ -t két részre bontva sorbafejtjük

$$N(N-1)\rho(z)(1-\mu(z)) = N(N-1)e^{-2z} = e^{-2y} \left(1 - \frac{1}{N} \right), \quad (3.2.39)$$

$$\mu(z)^{N-2} = \frac{e^{N \ln \left(1 - \frac{e^{-y}}{N} \right)}}{\left(1 - \frac{e^{-y}}{N} \right)^2} \approx e^{-e^{-y}} \left(1 - \frac{e^{-2y}}{2N} \right) \left(1 + \frac{2e^{-y}}{N} \right), \quad (3.2.40)$$

vagyis a sűrűségfüggvény $1/N$ szerinti sorfejtése

$$P_{(2)}^{\max}(y + \ln N, N) = e^{-2y-e^{-y}} \left(1 + \frac{2e^{-y}}{N} - \frac{e^{-2y}}{2N} - \frac{1}{N} \right), \quad (3.2.41)$$

behelyettesítve a (3.2.19) integrálba a második maximum várható értéke

$$\langle z_2 \rangle = \gamma_E - 1 + \ln N + \frac{1}{2N}. \quad (3.2.42)$$

Ugyanakkor az első maximum várható értéke analóg módon számolva [47]

¹(3.2.3) alaposabb vizsgálata azt mutatja, hogy $P_{(2);N}(x)$ kifejezésében megjelenik egy $\mathcal{O}(1/N)$ tag is, ezért (3.2.35) csak $\gamma' \geq -1$ esetén helytálló.

$$\langle z_1 \rangle = \gamma_E + \ln N + \frac{1}{2N}, \quad (3.2.43)$$

vagyis az első és a második maximum távolságának várható értéke az őseloszlás eredeti skáláján megadva

$$\langle z_1 \rangle - \langle z_2 \rangle = 1. \quad (3.2.44)$$

Tiszta exponenciális őseloszlás esetén az $N \rightarrow \infty$ limesz során ez a skála nulladrendben nem változik. (2.6.20) kifejezéshez visszatérve azonban láthatjuk, hogy az a_N skálaparaméter elsőrendű korrekciója szerepet játszik, hiszen eszerint $a_N = 1 - 1/2N$. Az (1) sztenderdizációra $1/a_N$ -nel szorozva tudunk áttérni, vagyis az első és a második maximum távolsága az (1) sztenderdizációban

$$d_1 = 1 + \frac{1}{2N}. \quad (3.2.45)$$

A távolság a (2) sztenderdizációban $1/\sigma_1$, ahol az első maximum szórása [47]

$$\sigma_1 = \frac{\pi}{\sqrt{6}} \left(1 - \frac{1}{2N} \frac{6}{\pi^2} \right), \quad (3.2.46)$$

ezzel a távolság a (2) sztenderdizációban

$$d_1^{\text{var}} = \frac{\sqrt{6}}{\pi} \left(1 + \frac{1}{2N} \frac{6}{\pi^2} \right). \quad (3.2.47)$$

Általános formában $\gamma = 0$ és $\gamma' = -1$ esetén

$$d_1^{\text{var}} = \frac{\sqrt{6}}{\pi} \left(1 + \epsilon_N \frac{6}{\pi^2} \right). \quad (3.2.48)$$

Összefoglalva azt mondhatjuk, hogy az FTG osztályban ($\gamma = 0$) az első és második legnagyobb érték távolságának (d_1) $N \rightarrow \infty$ határértéke a gyakorlati szempontból érdekes (2) sztenderdizációban $\sqrt{6}/\pi$. A távolság korrekciója véges méretű minta mellett $-\epsilon_N \frac{6}{\pi^2} \zeta(3)$ ha $\gamma' = 0$ (általánosított exponenciális, pl. Rayleigh-őseloszlás) és $\epsilon_N \frac{6}{\pi^2}$ ha $\gamma' = -1$ (tisztá exponenciális őseloszlás), ahol ϵ_N -t az őseloszlás aszimptotikus tulajdonságai határozzák meg az a_N és b_N transzformációs paramétereken keresztül (2.4.7) szerint.

4. fejezet

EXTRÉM ÉRTÉK STATISZTIKA VÉLETLEN MINTAMÉRET MELLETT

Az extrém értékek elméletének eddigi eredményei elsősorban azt az esetet írják le, amikor a minta mérete, amelyből a maximumot megfigyeljük nem egy véletlen érték. Jól ismert, hogy amennyiben ezzel a mintamérettel a végtelenhez tartunk, az iid változók extrém eloszlása a három határeloszlás egyikéhez fog tartani, hogy melyikhez, azt az őseloszlás nagy argumentumú viselkedése határozza meg. Ezekhez a határeloszlásokhoz tartozó véges méret korrekció is ismert elsőrendben [27], így az alkalmazások során tipikusan kis mintaméret esetét is kellő pontossággal tudjuk leírni.

Mivel elméleti szinten a fix N eset egyszerűbb, ezért az EVS alkalmazásakor a vizsgált adatsort azonos méretű intervallumokra, cellákra osztjuk, és a maximumokat ezeken a cellákon értelmezzük. Előfordulnak azonban olyan problémák, ahol a változó mennyiség térben vagy időben nem egyenletesen jelenik meg, ezért azonos méretű cellákra osztani körülményes. Jó példa erre az 5. fejezetben leírt probléma, ahol az égbolt adott térszögében megfigyelt galaxisok maximális fényességének eloszlását vizsgáljuk. Az éggömböt azonos térfogatú térszögekre osztjuk, amelyek a 2D-s vetületen azonos területű pixeleknek felelnek meg, azonban a galaxisok száma, vagyis a minta mérete, pixelről-pixelre változik. A maximális fényesség statisztikai leírásához ezért meg kell gondolnunk mennyiben változik a 'klasszikus' (fix N) EVS elmélete véletlen N mintaméret mellett.

Ebben a fejezetben iid változók extrém eloszlását vizsgáljuk véletlen N esetén, vagyis amikor a minta mérete, amelyen a maximumot megfigyeljük, egy véletlen változó, adott mintaméret eloszlással. A fix N EVS analógiájára, azt várjuk, hogy ha a mintaméret eloszlás egy tipikus N_0 értékével – például az átlaggal, legvalószínűbb értékkel – a végtelenhez tartunk, univer-

zális határeloszlások jelennek meg. Az, hogy véletlen N esetén a határeloszlás különbözik-e a fix N határeloszlástól, függ a mintaméret eloszlás $N_0 \rightarrow \infty$ határátmenet során mutatott skálázási tulajdonságaitól. Közös vonás, hogy a fix N mellett megismert univerzalizációs osztályok (Fisher-Tippett-Gumbel, Fisher-Tippett-Weibull, Fisher-Tippett-Fréchet osztály) változatlanul megmaradnak, s a random N határeloszlás függ egyrészt attól, hogy az őseloszlás melyik univerzalizációs osztályhoz tartozik, másrészt a mintaméret N_0 -lal skálázott függvényétől.

Bár a fizikusi köztudatba és az alkalmazási területekre még nem igazán kerültek be, a véletlen N esetén megjelenő EVS határeloszlások a matematikai irodalomban ismertek [46]. Mi fizikai szemszögből érthetőbb és könnyebben alkalmazható módon írjuk fel a határeloszlásokat, [46] eredményeivel összhangban, és a határeloszlásokon túl foglalkozunk a véges méret korrekció problémájával is. A 2. fejezetben ismertetett renormálási csoport megközelítés segítségével azt látjuk, hogy a vezető rendű FSC a fix N EVS analógiájára itt is felírható a transzformáció sajátfüggvényével.

4.1. Az univerzális osztályok megjelenése két speciális esetben

A következő analitikus megfontolások során – ahogy eddig is – az extrém érték statisztika alapproblémájából indulunk ki. Adott egy N elemszámú, (független, azonos eloszlású – iid) véletlen változókból álló minta, a cél a minta maximumának – $z = \max\{z_1, z_2, \dots, z_N\}$ – eloszlásának meghatározása.

A különbség itt a 'klasszikus' extrém érték elmélethez képest, hogy a minta mérete, N , nem egy fix paraméter a statisztika során, hanem z -hez hasonlóan szintén valószínűségi változó, adott eloszlással. Sűrűségfüggvényét (batch/sample size probability distribution) a továbbiakban $\hat{P}_{N_0}(N)$ ¹ jelöli, ahol N_0 a minta méretének egy tipikus értéke (átlaga vagy azzal szorosan összefüggő mennyiség).

Az N méretű cella megjelenésének valószínűsége tehát $\hat{P}_{N_0}(N)dN$, és egy N elemű cella maximumának eloszlása $P_N^{\max}(z)$, amit ismerünk az eredeti EVS elméletből. Ha $\rho(z)$ jelöli az őseloszlás sűrűségfüggvényét, míg $\mu(z) = \int_0^z \rho(z')dz'$ az integrált eloszlásfüggvényét, az N elemű minta maximumának

¹A $\hat{P}_{N_0}(N)$ eloszlás szigorúan véve egy diszkrét változó eloszlása, mégis folytonos eloszlással közelítjük, mivel a következő számolás során olyan határesetet veszünk ahol mind az eloszlás várható értéke, mind szórása a végtelenhez tart. A kontinuum limesz közelítés miatt azonban véges méret korrekció járuléka elveszhet.

eloszlása, a változók függetlensége esetén

$$M_N^{\max}(z) = \mu(z)^N, \quad (4.1.1)$$

$$P_N^{\max}(z) = \frac{d}{dz} M_N^{\max}(z) = N\mu(z)^{N-1}\rho(z). \quad (4.1.2)$$

Mivel a minta mérete 0 és ∞ között vehet fel értékeket, $\widehat{P}_{N_0}(N)dN$ valószínűséggel, az extrém eloszlást a következő integrál formában kaphatjuk meg

$$\widetilde{P}_{N_0}^{\max}(z) = \int_0^\infty \widehat{P}_{N_0}(N) P_N^{\max}(z) dN, \quad (4.1.3)$$

ami egy véges méretű extrém eloszlás, N_0 tipikus mintamérettel. Tételezzük fel, hogy – a 'klasszikus' EVS elmélethez hasonlóan – létezik olyan lineáris transzformáció, amellyel átranszformálva a változót ez az extrém eloszlás egy (nem degenerált) határeloszláshoz tart az $N_0 \rightarrow \infty$ határesetben

$$\widetilde{P}_{N_0}(x) = \widetilde{P}_{N_0}^{\max}(\widetilde{a}_{N_0}x + \widetilde{b}_{N_0}) \rightarrow \widetilde{P}(x). \quad (4.1.4)$$

A továbbiakban két speciális eset számolása során illusztráljuk, hogy létezik a fenti határeloszlás és alakja függ a mintaméret $\widehat{P}_{N_0}(N)$ eloszlásának alakjától. Egy korábbi eredményünk alapján, exponenciális őseloszlást véve, egyenletes illetve exponenciális $\widehat{P}_{N_0}(N)$ eloszlás esetén eltérő határeloszlásra jutottunk (a számolás részleteit illetően l. A. Függelék), annak ellenére hogy a mintaméret tipikus értékével, N_0 -lal tartottunk a végtelenhez, mely határesetben az exponenciális mintaméret eloszlás is az egyenletes eloszláshoz tart. Az őseloszlást tekintve ellenben találtunk univerzalitást a határeloszlások terén. Ugyanezt a számolást elvégezve Rayleigh-őseloszlásra, ugyanezeket a határeloszlásokat kapnánk, csupán a lineáris transzformáció paramétereit, \widetilde{a}_{N_0} és \widetilde{b}_{N_0} , értékei változnának. Ebből a megfigyelésből kiindulva itt egy általánosabb megközelítéssel fogjuk vizsgálni a problémát.

Az extrém eloszlás sűrűségfüggvényét adott őseloszlás és mintaméret eloszlás esetén (4.1.3) adja meg. A következő számolás szempontjából azonban egyszerűbb az integrált eloszlás használata

$$\widetilde{M}_{N_0}^{\max}(z) = \int_0^\infty \widehat{P}_{N_0}(N) M_N^{\max}(z) dN = \int_0^\infty \widehat{P}_{N_0}(N) \mu(z)^N dN, \quad (4.1.5)$$

(4.1.1) felhasználásával, ahol $\mu(z)$ az integrált őseloszlás. (4.1.4) limesszel analóg módon, tegyük fel hogy létezik az $N_0 \rightarrow \infty$ határesetben, egy megfelelő lineáris transzformációt végrehajtva, az integrált eloszlás határeloszlása

$$\widetilde{M}_{N_0}(x) = \widetilde{M}_{N_0}^{\max}(\widetilde{a}_{N_0}x + \widetilde{b}_{N_0}) \rightarrow \widetilde{M}(x). \quad (4.1.6)$$

Az univerzális vonások bemutatása érdekében az őseloszlás alakját nem specifikáljuk, csak a mintaméret eloszlását.

4.1.1. Egyenletes mintaméret eloszlás

Először tekintsük az egyenletes mintaméret eloszlást, vagyis $\hat{P}_{N_0}(N) = \frac{1}{N_0}$, ha $0 < N < N_0$. Ebben az esetben könnyen kiszámolható az egzakt extrém határeloszlás.

(4.1.5) alapján az integrált extrém eloszlás a következőképpen néz ki általános őseloszlásra

$$\widetilde{M}_{N_0}^{\max}(z) = \frac{1}{N_0} \int_0^{N_0} \mu(z)^N dN = \frac{1}{N_0} \frac{\mu(z)^N}{\ln \mu(z)} \Big|_0^{N_0} = \frac{\mu(z)^{N_0} - 1}{\ln \mu(z)^{N_0}}, \quad (4.1.7)$$

és az extrém eloszlás sűrűségfüggvénye

$$\widetilde{P}_{N_0}^{\max}(z) = \frac{\rho(z)}{\mu(z) \ln^2 \mu(z)} \left(\mu(z)^{N_0} \ln \mu(z) - \frac{1}{N_0} (\mu(z)^{N_0} - 1) \right). \quad (4.1.8)$$

Vegyük észre (4.1.7) kifejezésben, hogy az extrém eloszlás csak $\mu(z)^{N_0}$ -tól függ, ami nem más (4.1.1) szerint, mint a fix N_0 elemű minta maximumának eloszlása $M_{N_0}^{\max}(z)$, vagyis

$$\widetilde{M}_{N_0}^{\max}(z) = \frac{M_{N_0}^{\max}(z) - 1}{\ln M_{N_0}^{\max}(z)}. \quad (4.1.9)$$

Mivel a klasszikus (fix N) extrém elmélet alapján ismerjük az N_0 elemű minta maximumának határeloszlását (2.1.4)

$$M_{N_0}(x) = M_{N_0}^{\max}(a_{N_0}x + b_{N_0}) \rightarrow M(x), \quad (4.1.10)$$

könnyen látható, hogy az $N_0 \rightarrow \infty$ limeszben, a $z = a_{N_0}x + b_{N_0}$ transzformáció után milyen határeloszláshoz tart az integrált eloszlás

$$\widetilde{M}_{N_0}(x) = \widetilde{M}_{N_0}^{\max}(a_{N_0}x + b_{N_0}) \rightarrow \widetilde{M}(x) = \frac{M(x) - 1}{\ln M(x)}, \quad (4.1.11)$$

ahol (2.1.6) szerint

$$M(x) = e^{-(1+\gamma x)^{-1/\gamma}}. \quad (4.1.12)$$

Az extrém határeloszlás egyenletes mintaméret eloszlás esetén tehát

$$\widetilde{M}(x) = (1 + \gamma x)^{1/\gamma} \left(1 - e^{-(1+\gamma x)^{-1/\gamma}} \right). \quad (4.1.13)$$

Ez a formula leírja az extrém határeloszlást tetszőleges őseloszlás esetére. A γ paraméter az egyesített extrém határeloszlás (GEV) paramétere, értékét az

őseloszlás nagy argumentumú viselkedése határozza meg, ahogy azt már korábban láttuk. Az Fisher-Tippett-Gumbel osztályra érvényes határeloszlást a $\gamma \rightarrow 0$ határesetben kapjuk

$$\widetilde{M}(x) = e^x \left(1 - e^{-e^{-x}}\right). \quad (4.1.14)$$

Az extrém határeloszlás sűrűségfüggvényét egyenletes mintaméret eloszlás és általános őseloszlás esetén (4.1.13) deriválásával kapjuk. A három határeloszlás osztályra egyesített alakja

$$\tilde{P}(x) = \frac{1}{1 + \gamma x} \left[(1 + \gamma x)^{1/\gamma} \left(1 - e^{-(1+\gamma x)^{-1/\gamma}}\right) - e^{-(1+\gamma x)^{-1/\gamma}} \right], \quad (4.1.15)$$

az FTG osztályra, $\gamma \rightarrow 0$ határesetben

$$\tilde{P}(x) = e^x - e^{x-e^{-x}} - e^{-e^{-x}}, \quad (4.1.16)$$

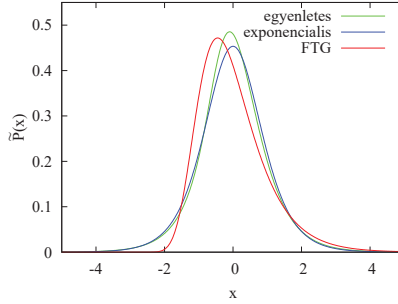
ami összhangban van az A. Függelékben az exponenciális őseloszlásra kapott eredménnyel (A.1.19), eltekintve az eltérő sztenderdizációs konstansoktól. (4.1.10) limesz és (4.1.12) formula alapján láthatjuk, hogy a fenti számolás során a fix N határeloszlást sztenderdizáltuk az (1) módszer szerint, és így kaptuk a legegyszerűbb alakban a véletlen N határeloszlást. A (2) sztenderdizációnak megfelelő eloszlást (A.1.19) alapján az $x \rightarrow a_0 x + b_0$ változócserevel kapjuk, ahol $a_0 = \sqrt{1 + \pi^2/6}$ és $b_0 = \gamma_E - 1$, és annak alakját a 4.1.1. ábrán láthatjuk.

4.1.2. Exponenciális mintaméret eloszlás

Analóg módon meghatározhatjuk az extrém határeloszlást exponenciális mintaméret eloszlás esetén, vagyis amikor a minta mérete, amelyből a maximumot húzzuk, $\hat{P}_{N_0}(N) = \frac{1}{N_0} e^{-N/N_0}$ eloszlást követő véletlen változó. Általános őseloszlás mellett az integrált extrém eloszlás (4.1.5) szerint

$$\widetilde{M}_{N_0}^{\max}(z) = \frac{1}{N_0} \int_0^\infty e^{-\frac{N}{N_0} \mu(z)} \mu(z)^N dN = \frac{1}{N_0} \left. \frac{e^{-\frac{N}{N_0} + N \ln \mu(z)}}{-\frac{1}{N_0} + \ln \mu(z)} \right|_0^\infty = \frac{1}{1 - \ln \mu(z)^{N_0}}, \quad (4.1.17)$$

mivel $\mu(z)$ egy integrált eloszlás, vagyis mindenhol kisebb, mint 1, így $\ln \mu(z) < 0$, ezért az integrál a felső határon eltűnik. Ebből az extrém eloszlás sűrűségfüggvénye



4.1.1. ábra. Extrém határeloszlások véletlen mintaméretre, egyenletes és exponenciális mintaméret eloszlás esetén, az FTG osztályban, a (2) sztenderdizációban. Ha a minta mérete N is véletlen változó $\hat{P}_{N_0}(N)$ sűrűségfüggvénnyel (ahol N_0 egy tipikus érték), a maximum határeloszlása az $N_0 \rightarrow \infty$ limeszben eltér a fix N határeloszlástól (FTG), amint azt két mintaméret eloszlás példáján láthatjuk, az egyenletes (A.1.19), és az exponenciális mintaméret eloszlás esetén (A.2.13).

$$\tilde{P}_{N_0}^{\max}(z) = \frac{\rho(z)}{\mu(z)} \frac{N_0}{(1 - \ln \mu(z)^{N_0})^2}. \quad (4.1.18)$$

Az egyenletes mintaméret eloszlás esetéhez hasonló módon az extrém eloszlás ismét csak $\mu(z)^{N_0} = M_{N_0}^{\max}(z)$ -től függ, vagyis másképp írva

$$\tilde{M}_{N_0}^{\max}(z) = \frac{1}{1 - \ln M_{N_0}^{\max}(z)}, \quad (4.1.19)$$

ezért (4.1.10)-t figyelembe véve az $N_0 \rightarrow \infty$ limeszben

$$\tilde{M}_{N_0}(x) = \tilde{M}_{N_0}^{\max}(a_{N_0}x + b_{N_0}) \rightarrow \tilde{M}(x) = \frac{1}{1 - \ln M(x)}. \quad (4.1.20)$$

Az exponenciális mintaméret eloszlás mellett kapott egyesített extrém határeloszlás és sűrűségfüggvénye tehát (4.1.12) felhasználásával

$$\tilde{M}(x) = \frac{1}{1 + (1 + \gamma x)^{-1/\gamma}}, \quad (4.1.21)$$

$$\tilde{P}(x) = \frac{(1 + \gamma x)^{-1/\gamma-1}}{(1 + (1 + \gamma x)^{-1/\gamma})^2}, \quad (4.1.22)$$

ahol γ -t az őseloszlás aszimptotikája határozza meg az extrém érték elméletből ismert módon. Az FTG osztályra, vagyis $\gamma \rightarrow 0$ esetre

$$\widetilde{M}(x) = \frac{1}{1 + e^{-x}}, \quad (4.1.23)$$

$$\widetilde{P}(x) = \frac{e^{-x}}{(1 + e^{-x})^2}, \quad (4.1.24)$$

ami megegyezik az exponenciális őseloszlás speciális esetére kiszámolt határeloszlás (A.2.13) szerinti alakjával. A (2) sztenderdizációra ennek megfelelően az $x \rightarrow \pi x/\sqrt{3}$ átskálázással térhetünk át, a határeloszlás alakját a 4.1.1. ábra mutatja.

A fenti két példa során láthattuk, hogy az extrém határeloszlás alakja függ a mintaméret eloszlástól ($\widehat{P}_{N_0}(N)$), az őseloszlások terén viszont univerzalitást mutat, az őseloszlás alakjától nem, csupán annak aszimptotikája által meghatározott γ paramétertől függ, vagyis adott $M(x)$ határeloszláshoz (FTF, FTG vagy FTW) $\widehat{P}_{N_0}(N)$ ismeretében kiszámolható a hozzá tartozó $\widetilde{M}(x)$. A 'klasszikus', fix N mintaméretű extrém érték statisztika során megismert univerzalitási osztályok tehát érvényben maradnak a változó mintaméretű EVS esetén is.

4.2. Extrém határeloszlás és véges méret korrekció általános esetben

4.2.1. Skálázó mintaméret eloszlás

A változó mintaméret mellett előálló extrém határeloszlás meghatározása tovább általánosítható, adott megkötéseknek eleget tevő mintaméret eloszlás esetén. Vegyünk először egy általános $\widehat{P}_{N_0}(N)$ mintaméret eloszlást. Az extrém eloszlás változó mintaméretre (4.1.5) alapján írható fel az őseloszlás ($\mu(z)$) ismeretében

$$\widetilde{M}_{N_0}^{\max}(z) = \int_0^\infty \widehat{P}_{N_0}(N) M_N^{\max}(z) dN = \int_0^\infty \widehat{P}_{N_0}(N) \mu(z)^N dN. \quad (4.2.1)$$

Tegyük fel, hogy amennyiben a tipikus mintamérettel végtelenhez tartunk ($N_0 \rightarrow \infty$) létezik egy (N_0 -tól független) extrém határeloszlás ($\widetilde{M}(x)$), amihez a véges méretű extrém eloszlás egy lineáris transzformáció végrehajtása után tartani fog

$$\widetilde{M}_{N_0}(x) = \widetilde{M}_{N_0}^{\max}(a_{N_0}x + b_{N_0}) \rightarrow \widetilde{M}(x). \quad (4.2.2)$$

Tegyük fel továbbá, hogy a mintaméret eloszlása a limesz végrehajtása során nem változtatja alakját, N_0 -al skálázó függvény

$$\widehat{P}_{N_0}(N)dN = \phi\left(\frac{N}{N_0}\right)\frac{dN}{N_0} = \phi(y)dy, \quad (4.2.3)$$

vagyis az eloszlás alakja nem, csupán várható értéke, és szórása függhet N_0 -tól. A véges méretű extrém eloszlás ekkor a következőképpen írható

$$\begin{aligned} \widetilde{M}_{N_0}(x) &= \int_0^\infty \widehat{P}_{N_0}(N)\mu(a_{N_0}x + b_{N_0})^N dN = \\ &= \int_0^\infty \phi(y)\mu(a_{N_0}x + b_{N_0})^{N_0 y} dy = \int_0^\infty \phi(y)[M_{N_0}(x)]^y dy, \end{aligned} \quad (4.2.4)$$

mivel (2.1.4) alapján

$$\mu(a_{N_0}x + b_{N_0})^{N_0} = M_{N_0}^{\max}(a_{N_0}x + b_{N_0}) = M_{N_0}(x) \rightarrow M(x; \gamma), \quad (4.2.5)$$

$N_0 \rightarrow \infty$ -re, az extrém határeloszlás változó mintaméret mellett

$$\widetilde{M}(x; \gamma) = \int_0^\infty \phi(y)M(x; \gamma)^y dy, \quad (4.2.6)$$

ahol $M(x; \gamma) = e^{-(1+\gamma x)^{-1/\gamma}}$ a 'klasszikus' fix N esetén megismert általános extrém határeloszlás (2.1.6). Vegyük észre, hogy a határeset végrehajtása során használt lineáris transzformáció paraméterei a_{N_0} és b_{N_0} az $M_{N_0}(x)$ eloszlást és ezáltal $M(x)$ határeloszlást sztenderdizálják az (1) módszer feltételei szerint. Ennek a transzformációnak köszönhetően $\widetilde{M}_{N_0}(x)$ nem divergál az $N_0 \rightarrow \infty$ limeszben, és a határeset egy nem degenerált $\widetilde{M}(x)$ határeloszlásra vezet. Bár $\widetilde{M}(x)$ (4.2.6) formulája nem felel meg az eddig használt sztenderdizációknak, formailag ebben az alakban a legegyszerűbb, és sztenderdizálása csak $\phi(y)$ skálafüggvény ismeretében lehetséges, ezért a továbbiakban is az elméleti megfontolások során így fogjuk használni. Amennyiben az $\widetilde{M}(x)$ határeloszlást empirikus adatokkal szeretnénk összevetni, az alkalmazások terén lényeges (2) sztenderdizációra a 2.5. fejezetben leírtak alapján tudunk áttérni.

Véletlen mintaméret mellett, ahol a mintaméret eloszlása $\phi(y)$ skálafüggvénnyel jellemezhető, és a tipikus mintamérettel $N_0 \rightarrow \infty$ -hez tartunk, az extrém eloszlás tehát a következő határeloszláshoz tart

$$\widetilde{M}(x; \gamma) = \int_0^\infty e^{-y(1+\gamma x)^{-1/\gamma}} \phi(y) dy, \quad (4.2.7)$$

$$\widetilde{P}(x; \gamma) = \int_0^\infty y(1+\gamma x)^{-1/\gamma-1} e^{-y(1+\gamma x)^{-1/\gamma}} \phi(y) dy, \quad (4.2.8)$$

ahol $1 + \gamma x \geq 0$ és γ értékét az őseloszlás aszimptotikus viselkedése határozza meg (l. 2.4. fejezet). Az FTG osztály esetén a határeloszlást a $\gamma \rightarrow 0$ limit elvégzésével kapjuk, a következő formában

$$\widetilde{M}(x; 0) = \int_0^\infty e^{-ye^{-x}} \phi(y) dy, \quad (4.2.9)$$

$$\widetilde{P}(x; 0) = \int_0^\infty ye^{-x-ye^{-x}} \phi(y) dy. \quad (4.2.10)$$

Átlátható és egyszerű alakban írható a véletlen mintaméret melletti határeloszlás a $\phi(y)$ skálafüggvény momentum generátor függvényének használatával

$$G_{\phi(y)}(s) = \langle e^{-sy} \rangle_{\phi(y)} = \int_0^\infty \phi(y) e^{-sy} dy, \quad (4.2.11)$$

mivel azt felhasználva (4.2.7) illetve (4.2.9) a következőképpen írható

$$\widetilde{M}(x; \gamma) = G_{\phi(y)}[(1 + \gamma x)^{-1/\gamma}], \quad (4.2.12)$$

$$\widetilde{M}(x; 0) = G_{\phi(y)}(e^{-x}). \quad (4.2.13)$$

A véges méret korrekció függvénye is egyszerűen számolható skálázó mintaméret eloszlás esetén. A 'klasszikus' EVS renormálási csoport elmélete szerint a véges méretű extrém eloszlás felírható egy az RG transzformáció fixpontja körüli kis perturbációként (2.3.12) alapján

$$M_{N_0}(x) = M(x) + \epsilon_{N_0} P(x) \psi(x), \quad (4.2.14)$$

ahol $M(x)$ az integrált határeloszlás (az RG transzformáció fixpontja), $P(x)$ a sűrűségfüggvénye és $\psi(x)$ a transzformáció sajátfüggvénye (2.3.6). (4.2.4) alapján a véges méretű, random N extrém eloszlás

$$\widetilde{M}_{N_0}(x) = \int_0^\infty \phi(y) [M_{N_0}(x)]^y dy \quad (4.2.15)$$

csak $M_{N_0}(x)$ kifejezésben tartalmazza N_0 -t (azzal a korábbi feltevéssel, hogy a mintaméret eloszlás N_0 -al skálázó függvény (4.2.3)), ezért a véges méret hatása itt is értelmezhető a fixpont körüli perturbációként

$$\begin{aligned} \widetilde{M}_{N_0}(x) &= \int_0^\infty \phi(y) [M(x) + \epsilon_{N_0} P(x) \psi(x)]^y dy = \\ &= \int_0^\infty \phi(y) M(x)^y \left[1 + \epsilon_{N_0} y \frac{P(x)}{M(x)} \psi(x) \right] dy = \\ &= \widetilde{M}(x) + \epsilon_{N_0} \psi(x) \int_0^\infty \phi(y) y M(x)^{y-1} P(x) dy, \end{aligned} \quad (4.2.16)$$

vagyis a véges méretű extrém eloszlás

$$\widetilde{M}_{N_0}(x) = \widetilde{M}(x) + \epsilon_{N_0} \widetilde{M}_1(x), \quad (4.2.17)$$

ahol a vezető rendű FSC függvénye

$$\widetilde{M}_1(x) = \widetilde{P}(x)\psi(x), \quad (4.2.18)$$

ϵ_{N_0} és $\psi(x)$ a 'klasszikus' EVS RG elméletében megismertek szerint az őseloszlásból meghatározható (2.4. fejezet)². A sűrűségfüggvény véges méret korrekciója

$$\widetilde{P}_{N_0}(x) = \widetilde{P}(x) + \epsilon_{N_0} \widetilde{P}_1(x), \quad (4.2.19)$$

ahol

$$\widetilde{P}_1(x) = \widetilde{P}(x) \left[\frac{\widetilde{P}'(x)}{\widetilde{P}(x)} \psi(x) + \psi'(x) \right]. \quad (4.2.20)$$

A véges méret korrekció tehát formailag megegyezik a fix N esetben fennálló korrekcióval, azzal a különbséggel, hogy a véletlen mintaméretnek megfelelő határeloszlás szerepel benne.

4.2.2. Nem skálázó mintaméret eloszlás

A gyakorlati alkalmazások során azonban a mintaméret eloszlás gyakran nem skálázik pontosan (4.2.3) szerint, a $\widehat{P}_{N_0}(N)$ alakja változik az $N_0 \rightarrow \infty$ limesz végrehajtása során. Az extrém határeloszlás létezése azonban megfogalmazható egy kevésbé szigorú feltétel fennállása mellett is. Amennyiben $\widehat{P}_{N_0}(N)$ nem skálázik (4.2.3) szerint, de tart egy skálafüggvényhez

$$\widehat{P}_{N_0}(N) dN = \widehat{\phi}_{N_0} \left(\frac{N}{N_0} \right) \frac{dN}{N_0} \rightarrow \phi(y) dy, \quad (4.2.21)$$

ha $N_0 \rightarrow \infty$, akkor a határeloszlás, amit korábban (4.2.4) szerint számoltunk, az integrandusban felbontva két tagra a limesz tulajdonságai miatt a következőképpen alakul

$$\begin{aligned} \widetilde{M}(x) &= \lim_{N_0 \rightarrow \infty} \int_0^\infty \widehat{P}_{N_0}(N) \mu(a_{N_0}x + b_{N_0})^N dN = \int_0^\infty \left(\lim_{N_0 \rightarrow \infty} \mu(a_{N_0}x + b_{N_0})^N \right) \cdot \\ &\quad \cdot \left(\lim_{N_0 \rightarrow \infty} \widehat{P}_{N_0}(N) dN \right) = \int_0^\infty \lim_{N_0 \rightarrow \infty} [M_{N_0}(x)]^y \phi(y) dy \end{aligned} \quad (4.2.22)$$

Vagyis a határeloszlás alakja – a skálázó esethez képest – változatlan maradt

$$\widetilde{M}(x; \gamma) = \int_0^\infty \phi(y) M(x; \gamma)^y dy. \quad (4.2.23)$$

²(4.2.17) szigorúan véve csak a marginális esetben érvényes, mivel $\widehat{P}_{N_0}(N)$ diszkrét eloszlás folytonossal való közelítéséből következően további $\mathcal{O}(1/N)$ rendű korrekciók jelenhetnek meg.

A véges méret korrekció viszont változik, mivel a mintaméret eloszlás a skálafüggvénytől különbözik az $N_0 \rightarrow \infty$ limesz végrehajtása során. Megjelenik egy új korrekciós tag, ami a $\hat{P}_{N_0}(N)$ tökéletes skálázásához képest jelen lévő tagokból jön.

A pontosan skálázó és a határesetben skálázó mintaméret eloszlás mellett vannak olyan esetek, amikor a mintaméret eloszlása, $\hat{P}_{N_0}(N)$ az N_0 -l való átskálázás során rendre szűkül az $N_0 \rightarrow \infty$ limesz során. $\hat{P}_{N_0}(N)$ összeszűkülését úgy is értelmezhetjük, hogy relatív szórása $\sigma_N / \langle N \rangle \rightarrow 0$, ha $N_0 \rightarrow \infty$, és az N_0 -l átskálázott mintaméret eloszlás delta-függvényhez tart

$$\hat{\phi}_{N_0} \left(\frac{N}{N_0} \right) \frac{dN}{N_0} \rightarrow \phi(y) dy = \delta(y - 1) dy. \quad (4.2.24)$$

A fenti megfontolások alapján létezik határeloszlás az $N_0 \rightarrow \infty$ limeszben, amit (4.2.23) határoz meg, vagyis

$$\widetilde{M}(x; \gamma) = M(x; \gamma), \quad (4.2.25)$$

a határeloszlás megegyezik a fix N esetén megjelenő határeloszlással. Az N véletlenszerűsége tehát keskenyedő mintaméret eloszlás esetén nem változtatja meg az extrém határeloszlás alakját, csupán az ahhoz tartozó véges méret korrekciót.

A fentiek alapján tehát $\hat{P}_{N_0}(N)$ határártmenet során mutatott skálázási tulajdonságai határozzák meg, hogy az extrém határeloszlás különbözik-e a fix N határeloszlásoktól. Amennyiben az $N_0 \rightarrow \infty$ limeszben létezik egy nem degenerált mintaméret skálafüggvény, $\phi(y)$, úgy az extrém határeloszlás függ annak alakjától, és (4.2.23) formában írható. Pontos skálázás esetén a véges méret korrekció is meghatározható a fix N EVS-sel analóg módon (4.2.17) a marginális esetben. Ha a mintaméret skálafüggvénye összeszűkül és határesetben egy delta-függvényhez tart, akkor az extrém határeloszlás a fix N elmélettel megegyezik (4.2.25). Előfordulhat, hogy nem létezik skálafüggvény a limesz végrehajtása során, és a skálázott eloszlás valamiképpen szétfolyik. Ez az eset azonban szerteágazó problémákat vet fel, amikkel itt nem foglalkoztunk.

4.2.3. A változó mintaméret mint korrekció

Amint azt a keskenyedő mintaméret eloszlás példáján láttuk – (4.2.24) és (4.2.25) –, ha $\hat{P}_{N_0}(N)$ relatív szórása nullához tart, akkor a változó mintaméret mellett előálló extrém határeloszlás megegyezik a fix mintaméret esetén megismert határeloszlással

$$\widetilde{M}(x) = M(x). \quad (4.2.26)$$

Eszerint, ha adott egy olyan $\widehat{P}_{N_0}(N)$ eloszlás, amely relatív szórása kicsi, az $\widetilde{M}(x)$ határeloszlás közel van $M(x)$ -hez, és ilyenkor a relatív szórást használva kis paraméterként $\widetilde{M}(x)$ sorbafejthető $\sigma_{\text{rel}} = 0$ körül, a sorfejtés elsőrendű tagja pedig korrekcióként kezelhető $M(x)$ -hez

$$\widetilde{M}(x) \approx M(x) + \sigma_{\text{rel}} M_1^{\text{randN}}(x). \quad (4.2.27)$$

Ugyanakkor véletlen mintaméret mellett, véges N_0 tipikus érték esetén az extrém eloszlás (4.2.17) alapján elsőrendben, nagy N_0 értékekre

$$\widetilde{M}_{N_0}(x) = \widetilde{M}(x) + \epsilon_{N_0} \widetilde{M}_1(x), \quad (4.2.28)$$

ebben a kifejezésben (feltéve, hogy σ_{rel} kicsi) (4.2.27)-t felhasználva és csak a vezető rendű tagokat meghagyva

$$\widetilde{M}_{N_0}(x) \approx M(x) + \sigma_{\text{rel}} M_1^{\text{randN}}(x) + \epsilon_{N_0} M_1(x), \quad (4.2.29)$$

ahol $M_1(x)$ a fix N -hez tartozó FSC, mivel (4.2.18) elsőrendű korrekciós tagban $\widetilde{P}(x) \approx P(x)$, mivel a magasabb rendű korrekciókat elhanyagoljuk.

Azt mondhatjuk tehát, hogy a mintaméret véletlen voltából eredő különbség is kezelhető – a véges mérettel analóg módon – korrekcióként. A változó mintaméret mellett előálló extrém eloszlás tehát, amennyiben mindkét sorfejtés jogos (σ_{rel} kellően kicsi, és N_0 nagy), felírható a határeloszlás és a két, véges és véletlen méretből adódó korrekció összegeként.

4.3. Az első és második legnagyobb érték távolsága skálázó mintaméret eloszlásra

Ebben az alfejezetben röviden megvizsgáljuk mennyiben változtatja meg a véletlen mintaméret az első és második legnagyobb érték várható távolságát az $N_0 \rightarrow \infty$ határesetben, a fix N esetén kapott (3.2.12) eredményhez képest.

Tekintsünk egy skálázó mintaméret eloszlást $\phi(y)$ skálafüggvénnyel, ekkor az első maximum határeloszlása (4.2.6) szerint

$$\widetilde{M}_{(1)}(x) = \int_0^\infty M_{(1)}(x)^y \phi(y) dy, \quad (4.3.1)$$

$$\widetilde{P}_{(1)}(x) = \int_0^\infty y(1 + \gamma x)^{-1/\gamma-1} M_{(1)}(x)^y \phi(y) dy. \quad (4.3.2)$$

A második legnagyobb érték eloszlását (4.1.5)-höz visszatérve határozhatjuk meg

$$\begin{aligned}\widetilde{M}_{(2)}^{\max}(z, N_0) &= \int_0^\infty M_{(2)}^{\max}(z, N) \widehat{P}_{N_0}(N) dN = \int_0^\infty M_{(2)}^{\max}(z, N_0 y) \phi(y) dy = \\ &= \int_0^\infty \mu(z)^{N_0 y} \left[1 + N_0 y \left(\frac{1}{\mu(z)} - 1 \right) \right] \phi(y) dy,\end{aligned}\quad (4.3.3)$$

ahol a második legnagyobb érték véges méretű eloszlását, $M_{(2)}^{\max}(x, N)$ -t (3.2.4) formulával helyettesítettük. Az átskálázás után $\widetilde{M}_{(2)}(x, N_0)$ az $N_0 \rightarrow \infty$ limitben a következő határeloszláshoz tart, (3.2.1) és (3.2.5) felhasználásával³

$$\widetilde{M}_{(2)}(x) = \int_0^\infty M_{(1)}(x)^y \left[1 + y(1 + \gamma x)^{-1/\gamma} \right] \phi(y) dy. \quad (4.3.5)$$

(4.3.1) és (4.3.2) felhasználásával láthatjuk, hogy a fix mintaméret esetével (3.2.10) megegyező formulával írható fel a második maximum határeloszlása

$$\widetilde{M}_{(2)}(x) = \widetilde{M}_{(1)}(x) + (1 + \gamma x) \widetilde{P}_{(1)}(x), \quad (4.3.6)$$

és emiatt, amint azt a fix mintaméretnél (3.2.11) is láttuk $\langle x_2 \rangle$ számolása során

$$\widetilde{d}_1 = \langle \widetilde{x}_1 \rangle - \langle \widetilde{x}_2 \rangle = \gamma \langle \widetilde{x}_1 \rangle + 1. \quad (4.3.7)$$

Változó mintaméret mellett tehát formailag ugyanazt kaptuk, mint fix mintaméret esetén (3.2.12)-ben, azzal a különbséggel hogy itt az $\langle \widetilde{x}_1 \rangle$ átlag – a (4.3.2) eloszlás várható értéke – függ a $\phi(y)$ skálafüggvénytől. Az FTG osztályon belül az első és második legnagyobb érték várható távolsága $\widetilde{d}_1 = 1$.⁴ Az alkalmazások szempontjából érdekes (2) sztenderdizációban, amire az $x \rightarrow (x - \langle \widetilde{x}_1 \rangle) / \widetilde{\sigma}_1$ transzformációval térünk át, a távolság az FTG osztályra

$$\widetilde{d}_1^{\text{var}} = \frac{1}{\widetilde{\sigma}_1}, \quad (4.3.8)$$

ahol az első legnagyobb érték szórása, $\widetilde{\sigma}_1$ függ $\phi(y)$ skálafüggvénytől.

³Az FTG osztály esetén a $\gamma \rightarrow 0$ limittel kapjuk a második legnagyobb érték véletlen mintaméretű, $N_0 \rightarrow \infty$ határeloszlását

$$\begin{aligned}\widetilde{M}_{(2)}(x) &= \int_0^\infty e^{-ye^{-x}} (1 + ye^{-x}) \phi(y) dy, \\ \widetilde{P}_{(2)}(x) &= \int_0^\infty e^{-ye^{-x}} (y^2 e^{-2x}) \phi(y) dy.\end{aligned}\quad (4.3.4)$$

⁴ne feledjük, hogy ez nem felel meg az (1) sztenderdizációnak

5. fejezet

GALAXISOK FÉNYESSÉGÉNEK EXTRÉM STATISZTIKÁJA AZ SDSS ADATAI ALAPJÁN

Az extrém érték statisztika gyakorlati alkalmazását nehezíti a határeloszlásokhoz való lassú konvergencia. Bár a véges méretből adódó korrekció elsőrendben ismert [26, 27], az FTG osztályban gyakori logaritmikus konvergencia miatt így is csak nagy adatbázisok esetén érhető el kellő egyezés az elmélettel. A Sloan Digital Sky Survey (SDSS) [50] mérései nyomán nagyszámú galaxisból álló fényesség adatbázis áll rendelkezésre, ami ezáltal alkalmas egy ilyen jellegű analízisre.

Ugyanakkor az extrém érték statisztika egy jó eszközt nyújthat az őseloszlások nagy argumentumú viselkedésének vizsgálatára, mivel mind a végtelen mintaméret határesetében megjelenő extrém határeloszlás, mind a véges méret korrekció érzékenyen függ az őseloszlás aszimptotikus viselkedésétől [27]. A galaxis fényességek extrém statisztikájának leírása így azért is érdekes lehet, mivel a galaxis térképezések egyik leggyakoribb statisztikai kérdése a luminozítás függvény meghatározása, ami ebben az alkalmazásban az őseloszlásnak felel meg. A luminozítás függvény alakja az elméleti Press-Schechter formula által meghatározott Schechter-függvény [62], egy hatványfüggvény exponenciális lecsengéssel. Az elmélet alapján azt várjuk, hogy az exponenciális lecsengésnek megfelelően a határeloszlás a Fisher-Tippett-Gumbel eloszlás, a véges méret korrekció amplitúdóját pedig a hatványkitevők határozzák meg. Az empirikus eloszlásokkal való összevetés így a luminozítás függvény exponenseinek meghatározásában segíthet. A véges méret korrekció, mint látni fogjuk, a nagyszámú adat ellenére is lényeges, és az EVS szempontjából elengedhetetlen.

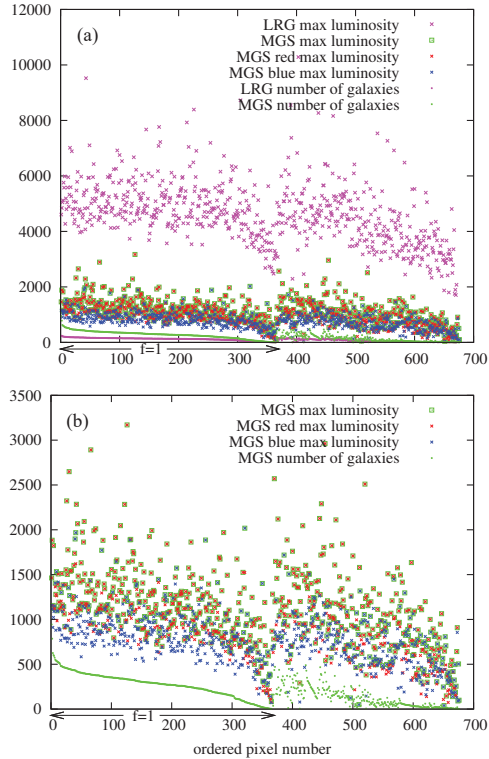
Egy adott térszögben vizsgáltuk a megfigyelt maximális galaxis fényesség eloszlását, ezért egy, az alkalmazások során felmerülő probléma vizsgálatát is szükségessé vált, a véletlen mintaméret hatásának vizsgálata. Az egyes irányokban eltérő számú galaxis látható, így a mintaelemek száma, amelyből a maximumot figyeljük meg egy – adott eloszlással rendelkező – változó. Emiatt az empirikus hisztogramok és az elméleti határeloszlások összevetésénél kétféle korrekciót is figyelembe veszünk, a véges méretből adódó korrekciót (FSC) és a véletlen mintaméretből adódó korrekciót. Az empirikus eloszlások jó egyezése a korrekciókkal kiegészített aszimptotikus elmélettel arra utal, hogy a Schechter-függvény lecsengése nagy argumentumok esetén is exponenciális marad, és nincs levágás egy nagy luminozitás értékénél.

A galaxisokat két fő csoportra osztottuk a vizsgálat során szín és morfológia szerint, úgy mint MGS (Main Galaxy Sample) és LRG (Luminous Red Galaxies). Mivel e két csoport luminozitás valamint térbeli eloszlása jelentősen eltér egymástól, ezért az extrém tárgyalás a csoportokra külön-külön lehet csak eredményes. Az MGS osztály szín szerint tovább bontható vörös és kék galaxisokra, melyek luminozitás eloszlása kissé eltér, de tárgyalásuk együttesen is lehetséges, így a fentiek alapján négy galaxis-csoportot vizsgáltunk: MGSall, MGSred, MGSblue és LRG-t.

Fontos momentum továbbá, hogy a galaxisok luminozitása (ami egyértelmű kapcsolatban áll a galaxisok tömegével) nem független változó. Az extrém érték statisztika azonban csak független és azonos eloszlású (iid) változókra ismert, a korrelált változók extrém elmélete csak pár speciális esetben tisztázott [31, 32]. Az iid elmélet azonban kiterjeszthető elegendően gyenge korreláció esetére is. A luminozitások közötti korrelációt csökkenti a mintavétel meghatározása, a maximumot egy térszögből választjuk, ami egy vékony és hosszú térfogategység, ún. pencil beam. Mivel a galaxisok tömege és ezáltal luminozitása között elsősorban rövid távú korrelációk vannak, egy ilyen jellegű mintaválasztás esetén jogos az iid feltételezés.

5.1. A használt adatbázis

Az égbolt több mint egynegyedének nagy mélységű exponálásával majd a világhálón elérhető adatbázissá való feldolgozásával (2000-2005 SDSS-I, 2005-2008 SDSS-II) a Sloan Digital Sky Survey napjaink legteljesebb asztronómiai katalógusává vált. A 230 millió égi objektumot tartalmazó nagyfelbontású színes képek elkészítése mellett, több mint 930 000 galaxisnak és 120 000 kvazárnak határozták meg a spektruma alapján vöröseltolódását, ezzel 3D-s térképet létrehozva.



5.1.1. ábra. Egy adott térszögben megfigyelt galaxisok száma és legnagyobb luminozitás értéke. A vízszintes tengelyen a térszögek indexe, amit rendeztünk az f arányszám, illetve a galaxis szám szerint. A vizsgálat során csak az $f \geq 0.97$ arányszámú adatokat használtuk, vagyis azokat a pixeleket, amelyek minimum 97%-át tartalmazza a feltérképezés. Az (a) ábrán jól látszik, hogy az LRG-k (magenta) fényesebbek, mint az MGS galaxisai (zöld), de átlagos számuk alacsonyabb. A (b) ábra az MGS populációt nagyítja ki, és annak további osztályozását mutatja be MGSred és MGSblue osztályokra. Jól látható, hogy a vörös galaxisok tipikusan fényesebbek, ezért az MGS all populáció maximumai között nagyobb arányban szerepelnek az MGSred galaxisai.

Az extrém analízishez galaxisok maximális luminozitásának statisztikáját vizsgáltuk, a használt adatok az SDSS DR8 fotometrikus és spektroszkopikus adatbázisából származnak [49, 50, 51], mely online hozzáférhető a *CasJobs* oldalán keresztül [52], és integrált kozmológiai függvények használatával elemezhető [53]. Az ég $F_A \approx 0.1923$ törtrészt felölelő spektroszkopiai adatbázisból azokat a galaxisokat használtuk, amelyek vöröseltolódásának mért értékei magas konfidenciaszinttel rendelkeztek a különböző mérési metódusok szerint. Kiszekeltálásra kerültek továbbá a hamis vagy a bizonytalan luminozitás értékek (pixelhatáron detektált galaxis, túl fényes, szaturált vagy profil nélküli galaxisok).

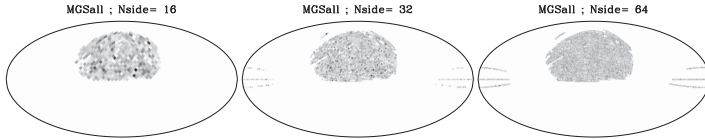
Azt tapasztaltuk, hogy a galaxisok összesített adatbázisának extrém vizsgálata nem lehetséges, mivel a szín, méret, morfológia és vöröseltolódás szerint kategorizált osztályokra jelentősen eltérhet mind a luminozitás eloszlása és átlagos értéke mind a galaxisok térbeli eloszlása. Az adatbázisban szereplő galaxisokat ezek alapján két fő csoportra osztottuk. Az egyik az MGS (Main Galaxy Sample) [54], mely a közeli, $([z_1, z_2] = [0.065, 0.22])$ vöröseltolódás közötti galaxisokat tartalmazza. A másik a magasabb vöröseltolódás értékeken $([z_1, z_2] = [0.2, 0.38])$ megfigyelt, fényes, de alapjában véve vörös színű, öreg, korai típusú galaxisok csoportja, az LRG (Luminous Red Galaxies) [55]. A vizsgálatok során az MGS (a továbbiakban MGS all) csoportot szín alapján további két alcsoportra osztottuk, úgy mint MGSred és MGSblue [56], ezáltal az extrém vizsgálatot négy galaxis populáción végeztük el (LRG, MGSall, MGSred, MGSblue). A két fő csoportban – MGS és LRG – a galaxisok száma $N_T \approx 349000$ volt (0.559 Gpc^3 -) illetve $N_T \approx 52600$ (2.18 Gpc^3 térfogatban). A különböző vöröseltolódás értékeken megfigyelt fényességeken (k+e)-korrekciót [57] alkalmaztunk az összehasonlíthatóság érdekében.

Mint arra már korábban utaltunk, nem minden alkalmazás során szükséges azonos méretű mintákra osztani az adatbázist az EVS vizsgálatához, mivel a gyakorlatban az adatbázis sokszor hiányos az észlelhetőség limitált-ságából adódóan, vagy gyakran egy folytonos változó idő vagy térbeli megvalósulásait vizsgáljuk. Mivel a galaxisok adatbázisa sem teljes (a láthatóság limitje miatt), a cellákat úgy definiáljuk, hogy azok az univerzum azonos térfogatú egységei legyenek. Az egyes térfogat egységek eltérő számú galaxist tartalmaznak, így ez a mintavétel lehetőséget nyújt az EVS vizsgálatára változó mintaméret mellett. Az égbolt 2D-s vetületének azonos területű régiókra osztása egyszerűen kivitelezhető az ún. HEALPix cellák (Hierarchical Equal Area isoLatitude Pixelisation) bevezetésével [58, 59]. Ezzel a feltérképezett területet azonos térfogatú pencil beam-szerű hosszú, keskeny cellákra tudtuk osztani, így minimalizálva a galaxis luminozitások közötti korrelációt, hogy az iid elmélet alkalmazható legyen. A cellák méretét illetően, három külön-

HEALPix cellák						
N_{side}	$\sqrt{\Omega_{pix}}$	n_{sphere}	$n_F/n_{F,97}$			
			MGSall	MGSblue	MGSred	LRG
16	3.66°	3072	768/473	767/473	767/473	755/473
32	1.83°	12288	2659/2030	2644/2030	2653/2030	2591/2029
64	55.0'	49152	10017/8492	9937/8489	9991/8490	9461/8256

5.1.1. táblázat. A mintavételi cellák mérete és száma a HEALPix geometria használatával. Az N_{side} értékkel jellemezzük a felbontást, Ω_{pix} az adott felbontáshoz tartozó pixel területe térszögben megadva, n_{sphere} a teljes égbolt kitöltéséhez szükséges HEALPix cellák száma, n_F a galaxist tartalmazó cellák száma a feltérképezett területen (SDSS footprint) és $n_{F,97}$ azon cellák száma, melyeknek legalább 97%-át tartalmazza a footprint.

bőző felbontást használtunk: $N_{side} = 16, 32$ és 64, a pixelek területe ebben a sorrendben két érték között negyedére csökken. Hogy mekkora pixelméret tartozik az egyes felbontásokhoz azt az 5.1. táblázatban láthatjuk, ahol a pixelméret, $\sqrt{\Omega_{pix}}$, térszögben van megadva.



5.1.2. ábra. Az SDSS által feltérképezett terület felosztása HEALPix cellákra a különböző N_{side} értékek esetén (MGSall populáció). A szürke árnyalatai az adott cellában megfigyelt maximális luminozítás értéket jellemzik.

Az égbolt adott felbontású HEALPix celláinak és az SDSS footprint-jének – az SDSS által feltérképezett régiók összességének – metszete (5.1.2. ábra) adja azokat a mintavételi cellákat, amelyből a maximumot megfigyeljük. Minden cella jellemezhető egy f arányszámmal, ami a footprint által lefedett rész arányát fejezi ki a cella területéhez képest. Az EVS analízishez csak azokat a cellákat használtuk, amelyekre $f \geq 0.97$, és tartalmaznak galaxist.

A vizsgált N_{side} megválasztása két ellentétes szempont alapján történik. Minél kisebb cellákra osztjuk az eget, annál jobb statisztikát kapunk, simább

hisztogramokat, hiszen a mintavételi cellák száma ezzel nő, ugyanakkor viszont a cellák méretének csökkenésével csökken a cellában található galaxisok átlagos száma, amelyből a maximumot megfigyeljük, ezért a maximumok eloszlása távolabb lesz a várható határeloszlástól, mivel nagyobb véges méret effektus lép fel. Bár $N_{side} = 64$ esetén a hisztogramok simák, mivel körülbelül $n_{F,97} \approx 8500$ adat áll rendelkezésre, a galaxisok átlagos száma cellánként $\langle N \rangle \approx 5 - 35$ között változik az egyes populációkra, ami az EVS szempontjából nagyon kevés. Ugyanakkor $N_{side} = 16$ esetén a cellánkénti átlagos galaxis szám $\langle N \rangle \approx 80 - 600$, míg a hisztogramokhoz használható adatok száma csak $n_{F,97} \approx 500$ (az átlagos galaxiszámokat illetően l. később az 5.3.1. ábrát és az 5.3.1. táblázatot).

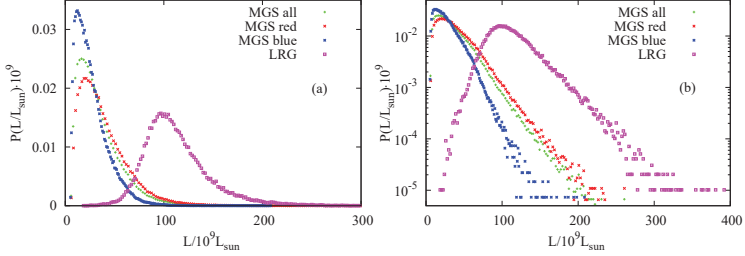
5.2. Luminozitás eloszlás

Az extrém statisztikai analízis szemszögéből rendkívül lényeges az őseloszlás ismerete, amelyből a maximumot megfigyeljük, különösen az aszimptotikus viselkedésének helyes leírása, hiszen ez határozza meg az extrém határeloszlás családot amelyhez a maximumok eloszlása tartani fog, valamint a véges méret korrekció alakját és amplitúdóját.

Az őseloszlás itt az SDSS által vizsgált térfogatban található galaxisok luminozításának eloszlása. A távcső szenzora által észlelt galaxisok luminozítás eloszlását azonban korrigálnunk kell, hogy megkapjuk a tényleges eloszlást, az asztrofizikában használatos ún. luminozítás függvényt. A luminozítás függvény megadja az univerzum egységnyi térfogatában lévő galaxisainak luminozítás eloszlását, vagyis az objektumok sűrűségét (számát) a luminozítás függvényében.

Az észlelt luminozítás eloszlásban a kevésbé fényes galaxisok kisebb súllyal szerepelnek (minél távolabb van egy objektum, annál nagyobb az az $L_{min}(r)$ érték, aminél kisebb fényességet a szenzor már nem képes érzékelni). A luminozítás függvényt az észlelt luminozítás eloszlás megfelelő súlyozásával kapjuk meg, ahol a különböző súlyok a térfogat-távolság-luminozítás összefüggéseket veszik figyelembe. Egy adott luminozítás érték felett ($M \leq -22.53$ az MGS-re és $M \leq -22.35$ az LRG-re) a mintánk teljes, vagyis ennél az értéknél fényesebb galaxisok mind láthatók az adott vöröseltolódás tartományban, és eloszlásukat nem kell súlyozni. Az észlelt eloszlás és a luminozítás függvény nagy argumentumú viselkedése ezért azonos. Az extrém statisztika szempontjából pedig csak a nagy argumentumú viselkedés érdekes. Ezért erre a közös aszimptotikára fittelt függvényt tekintjük az őseloszlás nagy argumentumú viselkedésének, ami az FTG osztályban meghatározza az extrém

viselkedést.



5.2.1. ábra. Az észlelt luminozitások eloszlása az SDSS r-színsávja alapján (a) lineáris és (b) féllogaritmikus skálán. Az MGS és az LRG galaxisaira a luminozitás eloszlás jelentősen eltér, mind alakját, mind várható értékét illetően.

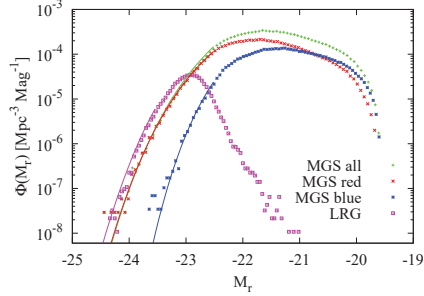
Az SDSS által észlelt galaxisok luminozitás eloszlását az 5.2.1. ábrán láthatjuk, valamint az asztrofizikai irodalom által használt magnitúdó eloszlást az 5.2.2. ábrán, ahol a magnitúdó és luminozitás közötti összefüggés [60]

$$M = -2.5 \log_{10}(L/L_{\odot}) + M_{\odot}, \quad (5.2.1)$$

és $M_{\odot} = 4.62$ az r-sávban. Amint az az ábrákon is látszik, a teljes galaxis adatbázis LRG és MGS osztályokra való bontása elengedhetetlen, mivel mind a luminozitás eloszlás alakja, mind várható értéke jelentősen eltér a két csoport esetén. Az MGS populációk luminozitás eloszlása, $P(L_r)$ exponenciálisan cseng le a nagy argumentumoknál (l. 5.2.1. ábra), a kis luminozitás értékekre pedig gyors hatványszerű lecsengéssel, majd $M_r \approx -19.6$ környékén levágva tart nullához, ezzel szemben az LRG-k eloszlása egy lognormális eloszláshoz hasonlít (l. 5.2.2. ábra). Az MGS galaxisait (MGSall a továbbiakban) tovább osztottuk szín szerint MGSred és MGSblue csoportra, melyek luminozitás eloszlása kissé eltér. Az észlelt luminozitás eloszlásokat megfelelő módon súlyozva meghatározhatjuk a luminozitás függvényeket [61], amiknek alakját az 5.2.3. ábrán láthatjuk az egyes populációkra.

Az SDSS adatainak elemzése szerint a Gamma-Schechter- vagy röviden Schechter-profil [62]

$$\Phi(L)dL = \Phi_* \left(\frac{L}{L_*} \right)^{\alpha} \exp \left(-\frac{L}{L_*} \right) \frac{dL}{L_*} \quad (5.2.2)$$



5.2.2. ábra. Az észlelt magnitúdó eloszlása az SDSS r-színsávja alapján a különböző galaxis-csoportokra féllogaritmikus skálán. Nagy negatív értékeknél a lecsengése jól leírható az általánosított Schechter-függvénnyel (5.2.3).

jól fitteli az MGS luminozitás függvényét. Korábbi vizsgálatok szerint a Schechter-profil α exponense a különböző színsávokban $\alpha = -1.35$ és $\alpha = -1.2$ között mozgott [60]. Mivel itt az extrém viselkedést vizsgáljuk, a Schechter-profil elsősorban az aszimptotika helyes leírását szem előtt tartva illesztettük. Az SDSS DR8 adatbázisának fittelése során kapott paramétereket az 5.2.1. táblázat tartalmazza.

A luminozitás függvény vizsgálata során a Schechter-profil mellett egy általánosabb függvényt is illesztettünk az MGS adataira, az általánosított Schechter függvényt (vagy általánosított gamma eloszlást), ami a Schechter-profil egy kiterjesztése

$$\Phi(L)dL = \Phi_* \left(\frac{L}{L_*} \right)^\alpha \exp \left[- \left(\frac{L}{L_*} \right)^\beta \right] \frac{dL}{L_*}, \quad (5.2.3)$$

és $\beta = 1$ rögzítése mellett az eredeti Schechter-profilat adja vissza. A luminozitás függvény aszimptotikájára ($M_r = -20.2$ felső limittel) illesztett paramétereket az 5.2.1. táblázat tartalmazza. Az MGS luminozitás függvénye mindkét fit (Schechter-, általánosított gamma eloszlás) szerint az FTG osztályba tartozik, mert lecsengése minden hatványnál gyorsabb. Az MGSall és MGSred esetében az általánosított gamma fit jobban leírja a nagy argumentumú viselkedést (l. 5.2.3. ábra), ezért az EVS során ezt használtuk öseloszlásként egységesen az MGS galaxisaira. Az észlelt és korrigáltan

	Φ_* [10^{-3}Mpc^{-3}]	M_*	α	β
Schechter profil				
MGSall	3.10 ± 0.05	-21.46 ± 0.02	-1.34 ± 0.04	1
MGSblue	2.70 ± 0.05	-20.87 ± 0.02	-1.06 ± 0.03	1
MGSred	2.06 ± 0.05	-21.41 ± 0.02	-0.99 ± 0.03	1
Általánosított gamma fit				
MGSall	7.79 ± 0.38	-20.42 ± 0.10	-0.81 ± 0.05	0.75 ± 0.02
MGSblue	1.79 ± 0.26	-21.22 ± 0.10	-1.28 ± 0.08	1.14 ± 0.05
MGSred	3.71 ± 0.11	-20.33 ± 0.17	-0.48 ± 0.08	0.74 ± 0.03
	Φ_* [10^{-5}Mpc^{-3}]	M_*	σ_M	γ
FTG fit				
LRG	2.52 ± 0.03	-22.85 ± 0.01	-21.36 ± 0.02	0
GEV fit				
LRG	2.49 ± 0.02	-22.86 ± 0.01	-21.36 ± 0.02	0.04 ± 0.01

5.2.1. táblázat. Luminozitás függvények illesztési paraméterei (magnitúdó értékekben megadva), Schechter-profil (5.2.2) és általánosított gamma eloszlás (5.2.3) illesztésével az MGS adataira, illetve FTG (5.2.4) és GEV (5.2.5) illesztésével az LRG-kra.

luminozitás eloszlásra illesztett általánosított gamma eloszlás paramétereit (5.2.2. táblázat) összevetve a luminozitás függvényével láthatjuk, hogy a nagy aszimptotikát meghatározó β paraméter közel azonos a két esetben. Az α paraméter ezzel szemben eltér, a luminozitás eloszlás fittelése során kapott pozitív hatványkitevő a kis luminozitás értékeknél írja le az eloszlás viselkedését. Ezzel egyidőben a nagy argumentumú viselkedés leírása pontatlanná válik, ezért az észlelt eloszlás fitje nem használható az EVS szempontjából.

Az LRG-k luminozitás függvényére extrém eloszlást illesztettünk, mivel ez a csoport egy adott redshift intervallum legfényesebb galaxisaiból áll (Luminous Red Galaxies), ezért ennek aszimptotikájára FTG-eloszlást illesztettünk

$$\Phi(L)dL = \Phi_* \exp \left\{ - \left(\frac{L - L_*}{\sigma} \right) - \exp \left[- \left(\frac{L - L_*}{\sigma} \right) \right] \right\} \frac{dL}{\sigma}, \quad (5.2.4)$$

$M_r = -22.64$ felső limittel. Az LRG populáció tehát úgy is értelmezhető, mint egy az MGS-hez hasonló őseloszlásból származó maximális luminozitások csoportja. A tényleges luminozitás függvényt azonban a legjobban egy

	M_*	α	β
Általánosított gamma fit			
MGSall	-19.99 ± 0.16	1.52 ± 0.10	0.79 ± 0.03
MGSblue	-21.02 ± 0.11	0.72 ± 0.10	1.18 ± 0.05
MGSred	-20.18 ± 0.16	1.67 ± 0.11	0.82 ± 0.03

5.2.2. táblázat. Luminozitás eloszlások illesztési paraméterei (magnitúdó értékekben megadva), általánosított gamma eloszlás (5.2.3) illesztésével az MGS adataira. Az illesztés során $\Phi_* = \beta/[L_*\Gamma((\alpha+1)/\beta, L_{\min})]$ rögzített, és $M_{\max} \equiv M(L_{\min}) = -20.2$ volt.

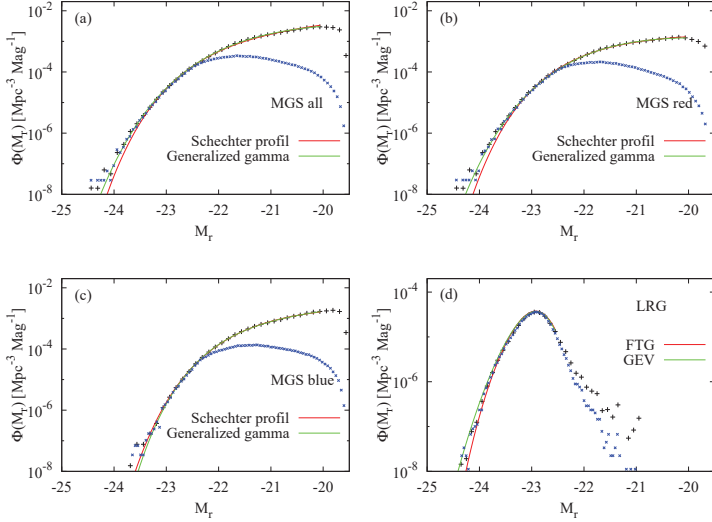
általánosított extrém eloszlás (GEV) írja le, mivel az LRG-k véges számú galaxis maximumai, vagyis az eloszlásuk egy véges méretű FTG-eloszlás, ami kis γ paraméterű GEV-nek felel meg [27]

$$\Phi(L)dL = \Phi_* \left[1 + \gamma \left(\frac{L - L_*}{\sigma_L} \right) \right]^{-1/\gamma-1} \exp \left\{ - \left[1 + \gamma \left(\frac{L - L_*}{\sigma_L} \right) \right]^{-1/\gamma} \right\} \frac{dL}{\sigma_L}. \quad (5.2.5)$$

Az LRG-k luminozitás függvénye jól fittelhető lognormális eloszlással is, ezt az őseloszlást azonban az EVS vizsgálatával zárhatjuk ki, mivel a határeloszláshoz való véges méret korrekciót nem magyarázza.

Az empirikus maximális luminozitás hisztogramok átlagos mintamérete többnyire viszonylag kicsi, így a véges méret korrekció lényeges. $N_{side} = 64$ esetén például – ahol a legkisebb a zaj, így leginkább kivethető az eloszlásfüggvény alakja – az átlagos mintaméret $\langle N \rangle = 5 - 35$ között mozog az egyes populációkra (5.3.1. táblázat). Ilyen kis mintaméretnél a véges méretű eloszlás határeloszlástól való eltérése $\approx 10\%$ a korrekció függvény középső szélsőértékénél (l. később az 5.4.5. ábrán).

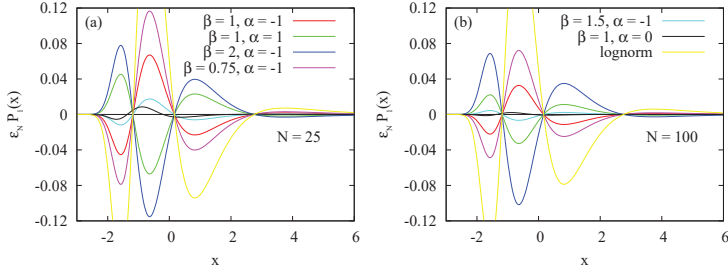
Az őseloszlás aszimptotikus viselkedésének helyes leírása azért is lényeges, mert meghatározza – a határeloszláson kívül – a véges méret korrekció (FSC) függvényének alakját, előjelét, amplitúdóját, konvergenciájának gyorsaságát (l. 2.4 fejezet). Általánosított gamma eloszlásra (aminek nagy argumentumú viselkedése megegyezik az általunk tárgyalt általánosított exponenciális aszimptotikával) az FSC alakját (2.6.14) szerint adtuk meg a (2) sztemerdizációban (nulla várható értékkel és egységnyi szórással), míg amplitúdóját (2.6.7) mindkét exponens értéke befolyásolja. Mivel ez a kifejezés az $N \rightarrow \infty$ határesetben érvényes, és N itt viszonylag kicsi, továbbá β nem sokkal tér el egytől, ezért az FSC amplitúdóját egy általánosabb, későbbiekben tárgyalt (5.4.11), mindkét tagot figyelembe vevő formában használjuk.



5.2.3. ábra. Luminozitás eloszlások (kék), luminozitás függvények (fekete) és az utóbbiak illesztései az SDSS DR8 adatai alapján az egyes galaxis csoportokra a magnitúdó térben ábrázolva. Az MGS luminozitás függvényei Schechter-profillal (5.2.2) és általánosított gamma eloszlással (5.2.3), míg az LRG FTG-eloszlással (5.2.4) és általános extrém eloszlással fittelve (5.2.5). Láthatjuk, hogy az észlelt luminozitás eloszlások aszimptotikája megegyezik a luminozitás függvényekével, tehát a fit-ek leírják mindkét függvény nagy argumentumú viselkedését.

Illusztrációként az 5.2.4. ábra mutatja néhány lehetséges őseloszlás-fit esetén fellépő véges méret korrekciót két különböző mintaméret mellett, hogy a különböző mértékű konvergenciát is megfigyelhessük. Az ábrán világosan látszik, mennyire érzékeny a véges méret korrekció függvénye az α és $\beta - 1$ előjelére. $\beta \approx 1$ esetén elsősorban α határozza meg a korrekció amplitúdóját (piros és zöld görbe), ami $1/\ln^2 N$ -tel cseng le. β értéke határozza meg az amplitúdót, ha az jelentősen eltér 1-től, ilyenkor a konvergencia lassabb, $1/\ln N$ -es (kék görbe). A két határeset között α és β együttesen határozzák meg ϵ_N -t (5.4.11) szerint, és a korrekció N -től függően akár előjelet is válthat (cián).

Az 5.2.4. ábra további két kivételes őseloszlás esetét is mutatja, amelyek



5.2.4. ábra. Véges méret korrekció néhány őseloszlás-fit esetén, és annak változása a mintamérettel, (a) $N = 25$ és (b) $N = 100$ mellett (a két ábra azonos színei ugyanazt az őseloszlást jelölik). Általánosított gamma eloszlás esetén az α ill. β exponens határozza meg az FSC-t (piros, zöld, kék, magenta, cían), tiszta exponenciális esetén gyors konvergenciára (fekete), lognormális őseloszlás esetén nagy korrekcióra (sárga) számíthatunk.

korábbi lehetséges fitjei az LRG luminozítás függvényének, a tiszta exponenciális esetén kicsi, $1/N$ -nel arányos, és a lognormális őseloszlás esetén nagy, $1/\sqrt{\ln N}$ -nel arányos korrekciót. Extrém határeloszlásból, amit itt az LRG-kra illesztettünk, mint őseloszlásból kiindulva nem lép fel véges méret korrekció, de ugyanakkor felléphet kis korrekció a változó mintaméret miatt, továbbá abból kifolyólag, hogy az illesztett hisztogram nem egzaktul FTG.

5.3. Galaxisok számának eloszlása

Az eget azonos térfogatú egységekre felosztó HEALPix cellákban megfigyelhető galaxisok száma celláról-cellára változik, vagyis a mintavételi cella mérete, N , amelyből a maximumot húzzuk egy változó mennyiség, adott $P(N)$ eloszlással. Az elméleti $P(N)$ eloszlást nem ismerjük, ezért az empirikus galaxis darabszámok alapján közelítjük egy a hisztogramra illesztett modell függvényrel.

Az empirikus galaxiszám eloszlásokat a különböző galaxis populációkra és különböző HEALPix felbontásokra az 5.3.1. összegző ábra mutatja. Láthatjuk, hogy az LRG és MGS minták nemcsak luminozítás de térbeli eloszlásukban is eltérnek. Szembetűnő, hogy a vörös galaxisok (MGSred) eloszlása általában lassabban cseng le, mint a kék galaxisoké (MGSblue), ennek ma-

	$N_{side} = 64$		$N_{side} = 32$		$N_{side} = 16$	
	$\langle N \rangle$	σ_{rel}	$\langle N \rangle$	σ_{rel}	$\langle N \rangle$	σ_{rel}
MGSall	36	0.443	142	0.333	567	0.249
MGSred	22	0.512	89	0.371	354	0.267
MGSblue	13	0.444	53	0.313	213	0.234
LRG	5	0.571	20	0.368	80	0.237

5.3.1. táblázat. A relatív szórás, $\sigma_{rel} = \sigma_N / \langle N \rangle$ változása $\langle N \rangle$ növelésével. $N_{side} = 64$ -től $N_{side} = 16$ felé a pixelenkénti átlagos galaxis szám, $\langle N \rangle$, növekszik (a pixelek egyre nagyobbak), ezzel együtt a relatív szórás csökken, vagyis a galaxis szám eloszlása, $P(N)$, relatíve egyre szűkebb.

gyarázata hogy a vörös galaxisok tipikusan nagyobb klasztereket alkotnak [63]. A különbségek ellenére azonban a négy galaxis populáció eloszlása egybeejtethető az átlaggal ($N \rightarrow N / \langle N \rangle$) vagy a szórással ($N \rightarrow (N - \langle N \rangle) / \sigma_N$) való átskálázás után. Mivel a változó mintaméret elmélete szerint egy skála paraméter lényeges, az N_0 (4. fejezet), az átlaggal való skálázást fogjuk használni, az átlagot választva tipikus mintaméretnek ($N_0 = \langle N \rangle$).

A $P(N)$ eloszlások statisztikai tulajdonságait nézve két megfigyelés is alátámasztja a klasszikus (fix N) EVS elmélet alkalmazhatóságát a maximális luminozitások terén. Egyrészt az N eloszlása viszonylag éles, relatív szórása $\sigma_N / \langle N \rangle = 0.2 - 0.5$ körül mozog. Másrészt az $\langle N \rangle$ növelésével (a cella méretének növelésével) a relatív szórás csökken az egyes populációkra (5.3.1. táblázat). Mint azt a 4. fejezetben láttuk, amennyiben a $P(N)$ eloszlás egy delta függvényhez tart (vagyis a relatív szórás $\sigma_N / \langle N \rangle \rightarrow 0$) ahogy $\langle N \rangle \rightarrow \infty$, az extrém határeloszlás marad a fix N elmélet szerinti, ami jelen esetben az FTG-eloszlás. Látna, hogy a relatív szórás szisztematikusan csökken az átlag növelésével, és feltesszük, hogy nullához tart, azt mondhatjuk, hogy az FTG-eloszlás a maximumok határeloszlása. Másrészt, a pontos leírás érdekében véges N esetén a változó N korrekciót is figyelembe kell vennünk az FSC mellett.

Ahhoz, hogy elméleti meggondolásokat tegyünk ill. hogy szimulációkat végezhessünk, érdemes egy jól fittelő függvénnyel modellezni a galaxiszám eloszlást. Az $N_{side} = 64$ felbontást tekintve az eloszlások aszimmetrikusak, és nagy N aszimptotikájuk exponenciálisan cseng le. Jól leírja az eloszlásfüggvények alakját a gamma eloszlás

$$P_1(x) = \frac{a}{\Gamma(k+1)} (ax - b)^k e^{-(ax-b)}, \quad (5.3.1)$$

($ax + b \geq 0$), amelyet $k = 3$ -mal illesztettünk. A skálaparaméter mellett

	skewness		
	$N_{side} = 64$	$N_{side} = 32$	$N_{side} = 16$
MGSall	1.566	1.285	0.685
MGSred	1.908	1.489	0.769
MGSblue	0.897	0.785	0.512
LRG	0.979	0.487	0.355

5.3.2. táblázat. A galaxisszám eloszlás, $P(N)$ ferdesége (skewness) csökken a felbontás csökkentésével, vagyis növekvő $\langle N \rangle$ -gal.

egy eltolási paraméterre is szükség volt, mivel kis N -eknél levágás van az eloszlásban az adatfeldolgozás során bevezetett f arányszám korlátja ($f \geq 0.97$) következményeként.

Vegyük észre, hogy az N_{side} csökkenésével, vagyis a pixelméret növelésével az eloszlás egyre szimmetrikusabb lesz, ferdesége (skewness) csökken, l. 5.3.2. táblázat, és az eloszlás egyre jobban fittelhető normális eloszlással

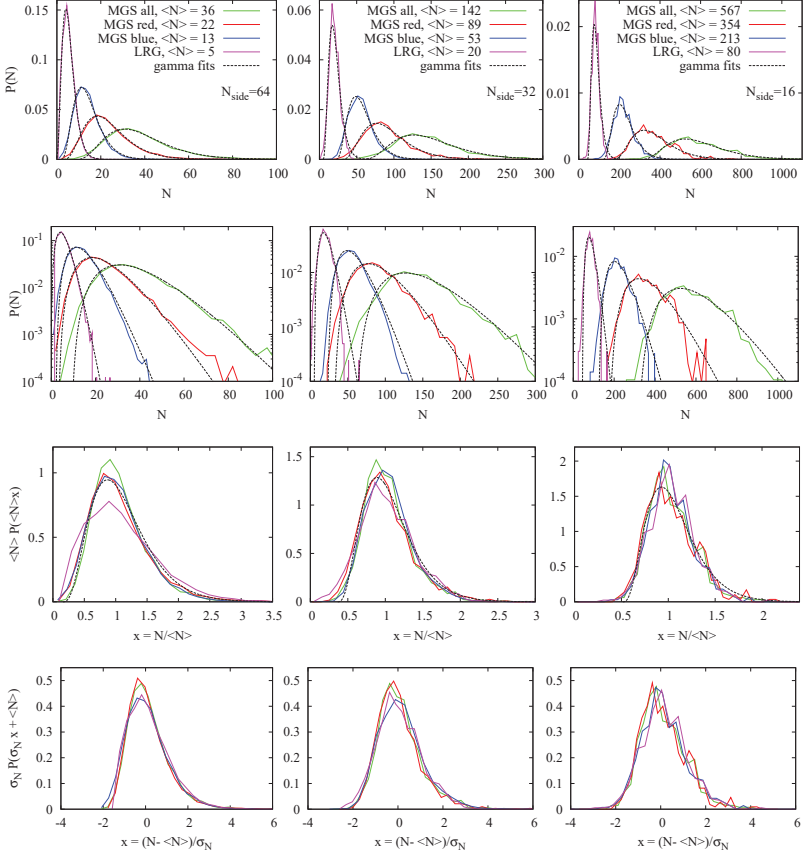
$$P_G(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-m)^2}{2\sigma^2}}. \quad (5.3.2)$$

A pixelméret növelése tulajdonképpen a kisebb pixelek egybeolvasztásával történik, tehát a galaxisszámok összeadásával, ezért az eloszlás Gausshoz tartása értelmezhető a centrális határeloszlás jegyében (feltéve, hogy a galaxisszámok legfeljebb gyengén korreláltak).¹

Mivel a gamma eloszlás mindhárom felbontás mellett jól illeszthető függvény, és az elméleti megfontolások is könnyen kivitelezhetőek egy ilyen mintaméret eloszlással, ezért ezt használtuk modellként. Az egységes leírás érdekében az átlaggal összeszkálázott mintaméret eloszlásra is meghatároztuk az illesztési paramétereket, és az elméleti eloszlások meghatározása során ezeket a paramétereket használtuk. A különböző felbontású HEALPix cellákban megfigyelhető galaxisszám eloszlásokra illesztett gamma eloszlásokat az 5.3.1. ábrán láthatjuk, az illesztési paramétereket az 5.3.3. táblázat tartalmazza.

¹ megjegyezzük, hogy ugyancsak jól modellezhető a galaxisszám eloszlás a negatív binomiális eloszlással is

$$P_{\text{nb}}(x) = \frac{\Gamma(x+r)}{\Gamma(x+1)\Gamma(r)} (1-p)^r p^x. \quad (5.3.3)$$



5.3.1. ábra. Galaxisok számának eloszlása egy HEALPix cellában, a négy különböző populációra (MGSall - zöld, MGSred - piros, MGSblue - kék, LRG - magenta). Balról jobbra $N_{\text{side}} = 64, 32, 16$, és az átlagos galaxiszám növekszik. Az eloszlásokat gamma eloszlással (5.3.1) illesztettük (1. és 2. sor lineáris és félogaritmikus skálán), az illesztés paramétereit az 5.3.3. táblázat tartalmazza. Az egyes populációk eloszlása egybeesik ha egységnyi átlagra (3. sor), vagy nulla átlagra és egységnyi szórásra (4. sor) skálázzuk a hisztogramokat.

	$N_{side} = 64$		$N_{side} = 32$		$N_{side} = 16$	
	a	b	a	b	a	b
MGSall	0.136	1.19	0.044	2.54	0.014	4.18
MGSred	0.194	0.57	0.065	1.95	0.020	3.31
MGSblue	0.323	0.61	0.113	2.49	0.037	4.42
LRG	0.674	0.26	0.240	1.13	0.091	3.88
$P(N/\langle N \rangle)$	4.220	0.67	5.734	2.09	7.276	3.75

5.3.3. táblázat. Gamma eloszlás (5.3.1) illesztési paraméterei a galaxiszám eloszlás leírására a négy populáció esetén. Az átlaggal összekészített négy populáció együttes illesztésének paramétereit az utolsó sor mutatja.

5.4. A maximális luminozítások elméleti határeloszlása és véges méret korrekciója

A maximális luminozítások eloszlását az egyes HEALPix cellákban megfigyelt galaxisok között mért legnagyobb fényesség értékekből készített hisztogrammal vizsgáltuk. Négy különböző galaxis populációra valamint három N_{side} felbontásra, így 12 hisztogramot vetettünk össze az elmélettel, melyekre a mintaméret tipikus értéke – ezáltal a véges méret korrekció – és a zaj mértéke eltért (l. később az 5.4.5. ábrán). Az eltérő tipikus mintaméret mellett felvett hisztogramokat az összehasonlíthatóság kedvéért nulla várható értékűre és egységnyi szórásra skáláztuk az $x = (L_{\max} - \langle L_{\max} \rangle) / \sigma$ változó bevezetésével. A hisztogramok a Fisher-Tippett-Gumbel-eloszlással összevetve jó egyezést mutattak, de ez az egyezés csak úgy értelmezhető, ha két ellentétes előjelű korrekció részleges kioltásából következik. Az empirikus eloszlások FTG-től való eltérése ugyanis a legtöbb esetben jóval kisebb a várható véges méret korrekciónál, ennek egy lehetséges magyarázata, hogy a véletlen mintaméretből adódó korrekció részben kioltja a véges méret korrekciót.

Ebben az alfejezetben analitikus módszerekkel próbálunk egy jó becslést adni a HEALPix cellánkénti maximális luminozítás eloszlására. Mivel a galaxisok száma cellánként eltér, a véletlen mintaméretű extrém statisztikák kell használnunk. Ahogy azt a 4.2. fejezetben láttuk, véletlen mintaméret mellett is analitikusan számolható mind az $N_0 \rightarrow \infty$ esetén érvényes extrém határeloszlás mind a véges méret korrekció elsőrendű tagja, amennyiben a mintaméret eloszlás skálázik N_0 -al. A határeloszlás alakja két tényezőtől függ, az őseloszlás aszimptotikus viselkedésétől (vagyis a γ paramétertől), valamint a mintaméret N_0 -al skálázott eloszlásától, $\phi(y)$ -től. Ezeket a tényezőket az empirikus adatokra való illesztésekből határozzuk meg. Jelen

esetben az őseloszlás a luminozítás függvény, amelynek lecsengése minden hatványnál gyorsabb, exponenciális, tehát $\gamma = 0$, az őseloszlás az FTG vonzási tartományába tartozik. A mintaméret eloszlása a galaxisszám eloszlás, ami a három általunk vizsgált felbontás esetén N_0 növelésével keskenyedett, relatív szórása csökkent. Amennyiben az $N_0 \rightarrow \infty$ határesetben $\sigma_{N_0}/N_0 \rightarrow 0$ teljesül, a 4.2. fejezetben vázoltak alapján a határeloszlás az FTG-eloszlás lesz (akár a fix N esetén). Azonban, ha a relatív szórás egy nem nulla értékhez tart, vagyis a skálázott mintaméret eloszlás, $\phi(y)$, nem egy delta-függvény az $N_0 \rightarrow \infty$ limeszben, a határeloszlás az FTG-től eltér, de $\phi(y)$ ismeretében kiszámolható.

Tekintsük mintaméret eloszlásként a gamma eloszlást, ami jól fitteli az empirikus galaxisszám eloszlásokat

$$\hat{P}_{N_0}(N) = \frac{k+1}{N_0 \Gamma(k+1)} \left(\frac{(k+1)N}{N_0} \right)^k e^{-\frac{(k+1)N}{N_0}}, \quad (5.4.1)$$

ahol $N_0 = \langle N \rangle$, a $\hat{P}_{N_0}(N)$ eloszlású mintaméret várható értéke. A skálázott mintaméret eloszlás ($y = N/N_0$ változócserevel)

$$\phi(y) = \frac{k+1}{\Gamma(k+1)} ((k+1)y)^k e^{-(k+1)y}. \quad (5.4.2)$$

Az extrém határeloszlás tehát (4.2.6) alapján

$$\widetilde{M}(x) = \frac{k+1}{\Gamma(k+1)} \int_0^\infty ((k+1)y)^k e^{-(k+1)y} M(x)^y dy, \quad (5.4.3)$$

ahol – mivel $\gamma = 0 - M(x)$ az FTG-eloszlás

$$M(x) = e^{-e^{-x}}. \quad (5.4.4)$$

Az elméleti extrém határeloszlás tehát, amihez az átlagos mintaméret növelésével ($N_0 \rightarrow \infty$) a maximális luminozítások eloszlása tartani fog

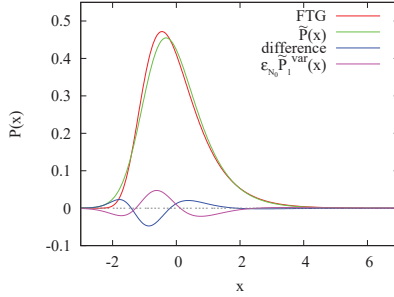
$$\widetilde{M}(x) = \frac{k+1}{\Gamma(k+1)} \int_0^\infty ((k+1)y)^k e^{-(k+1)y(1+e^{-x})} dy = \frac{1}{\left(1 + \frac{e^{-x}}{k+1}\right)^{k+1}}, \quad (5.4.5)$$

sűrűségfüggvénye pedig a következő

$$\widetilde{P}(x) = \frac{e^{-x}}{\left(1 + \frac{e^{-x}}{k+1}\right)^{k+2}}. \quad (5.4.6)$$

A határeloszlás alakját, és az FTG-eloszlástól való eltérését az 5.4.1. ábra mutatja, $k = 3$ esetén (az alkalmazások terén praktikusabb (2) típusú sztenzardizációban – egységnyi szórás, nulla várható érték – ami az FTG-eloszlás

esetén az $x \rightarrow \frac{\pi}{\sqrt{6}}x + \gamma$, $\tilde{P}(x)$ esetén az $x \rightarrow \sqrt{\frac{\pi^2}{3} - \frac{49}{36}}x + \frac{11}{6} - \ln(4)$ változó-cserével valósítható meg). Bár a sűrűségfüggvény alakja az FTG-hez nagyon hasonló, különbsége attól jelentős (10% körüli). Viszont adott esetben, amint azt az ábra is jól illusztrálja, a véletlen N hatása és a véges méret korrekció éppen kiolthatja egymást.



5.4.1. ábra. Ha a mintaméret eloszlása gamma eloszlás, $\phi(y) = \frac{1}{1.5}(4y)^3 e^{-4y}$ skálafüggvénnyel, az elméleti extrém határeloszlás, $\tilde{P}(x)$ alakja az FTG-eloszláséhoz hasonló, de különbségük jelentős. Mivel a különbségfüggvény maximum-minimum helyei a véges méret korrekcióival közel egybeesnek, és előjelük adott esetben lehet ellentétes, a véges méretű extrém eloszlás, $\tilde{P}_{N_0}(x) = \tilde{P}(x) + \epsilon_{N_0} \tilde{P}_1(x)$ az FTG-eloszláshoz nagyon hasonló lehet (az itt ábrázolt FSC görbe például $\epsilon_{N_0} = \frac{1-\beta}{\beta \ln N_0}$, ahol $\beta = 0.75$ és $N_0 = 68$ esetén áll elő).

Empirikus adatok vizsgálata során azonban mindig véges méretű eloszlásról beszélünk, ezért – és a leggyakrabban logaritmikusan lassú konvergencia miatt – figyelembe kell vennünk a véges méret korrekciót is, aminek alakja szintén analitikusan számolható. Elsőrendű tagja (4.2.18) szerint

$$\tilde{M}_1(x) = \tilde{P}(x)\psi(x), \quad (5.4.7)$$

ezzel a véges méretű eloszlás

$$\tilde{M}_{N_0}(x) = \tilde{M}(x) + \epsilon_{N_0} \tilde{M}_1(x) \quad (5.4.8)$$

alakban írható. A 2.6. fejezetben meghatároztuk a Schechter típusú (5.2.3) őseloszláshoz tartozó FSC paramétereket. Beláttuk, hogy az őseloszlásból $\gamma = \gamma' = 0$ következik, vagyis a véges méret korrekció sajátfüggvénye az (1) sztenderdizációban (2.3.9)

$$\psi_{0,0} = \frac{-x^2}{2}. \quad (5.4.9)$$

A véges méret korrekció nagyságrendje azonban függ a β exponens 1-hez való viszonyától (2.6.7). $\beta \not\approx 1$ mellett az α exponenstól független és $1/\ln N_0$ -al arányos az amplitúdó, míg $\beta \approx 1$ esetén függ α -tól, és a konvergencia menete gyorsabb

$$\epsilon_{N_0} \approx \begin{cases} \frac{1-\beta}{\beta \ln N_0}, & \beta \not\approx 1, \\ \frac{-\alpha}{\ln^2 N_0}, & \beta \approx 1. \end{cases} \quad (5.4.10)$$

Mivel az empirikus őseloszlások alapján β értéke egyhez közeli, és a vizsgált esetekben $N_0 \sim \langle N \rangle$ sem túl nagy, ezért egy általánosabb, mindkét tagot figyelembe vevő formában használjuk ϵ_{N_0} -t a korrekciók értelmezéséhez. (2.6.5) formulához visszatérve és figyelembe véve, hogy az őseloszlás alakja itt $\rho(z) \sim z^\alpha \exp(-z^\beta)$, vagyis a δ, θ exponensek definíciója alapján $\delta = \beta$ és $\theta = \beta - \alpha - 1$ -nek felel meg, ha $\beta > 0$, így

$$\epsilon_{N_0} = \frac{1-\beta}{\beta \ln N_0} + \frac{(2\beta-1)(\beta-\alpha-1)}{\beta^2 \ln^2 N_0}. \quad (5.4.11)$$

Az integrált eloszláshoz tartozó elsőrendű véges méret korrekció függvénye (5.4.6) határeloszlás és (5.4.9) sajátfüggvény felhasználásával

$$\widetilde{M}_1(x) = -\frac{1}{2} \frac{x^2 e^{-x}}{\left(1 + \frac{e^{-x}}{k+1}\right)^{k+2}}, \quad (5.4.12)$$

a sűrűségfüggvény korrekciója pedig (4.2.20) szerint

$$\tilde{P}_1(x) = \tilde{P}(x) \left[\frac{\tilde{P}'(x)}{\tilde{P}(x)} \psi(x) + \psi'(x) \right], \quad (5.4.13)$$

ahol (5.4.6) felhasználásával látjuk, hogy

$$\frac{\tilde{P}'(x)}{\tilde{P}(x)} = -1 + \frac{k+2}{1 + (k+1)e^x}, \quad (5.4.14)$$

és ezzel

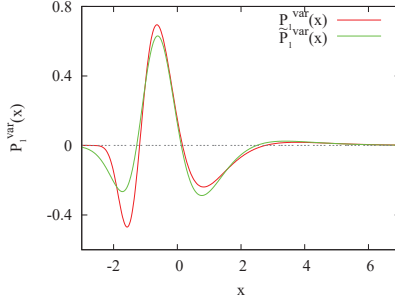
$$\tilde{P}_1(x) = \tilde{P}(x) \left(-x + \frac{x^2}{2} - \frac{(k+2)x^2}{2 + 2(k+1)e^x} \right). \quad (5.4.15)$$

Ahhoz, hogy a véges méret korrekció az empirikus adatokkal összevethető legyen, át kell térni a (2) típusú, vagyis variancia szerinti, sztenderdizációra, amit a sűrűségfüggvény esetén (2.5.9) alapján tehetünk meg.

$$\begin{aligned}
\tilde{P}_1^{\text{var}}(x) &= a_o \tilde{P}(a_o x + b_o) \left[\frac{\tilde{P}'(a_o x + b_o)}{\tilde{P}(a_o x + b_o)} \left(\psi(a_o x + b_o) + a_1 x + b_1 \right) + \right. \\
&\quad \left. + \psi'(a_o x + b_o) + \frac{a_1}{a_o} \right] = a_o \tilde{P}_1(a_o x + b_o) + a_o \tilde{P}(a_o x + b_o) \cdot \\
&\quad \cdot \left[\left(\frac{k+2}{1+(k+1)e^{a_o x + b_o}} - 1 \right) (a_1 x + b_1) + \frac{a_1}{a_o} \right] \quad (5.4.16)
\end{aligned}$$

ahol a paraméterek $\tilde{P}_{N_0}(x) = \tilde{P}(x) + \epsilon_{N_0} \tilde{P}_1(x)$ eloszlásból meghatározhatók (2.5.3) és (2.5.4) szerint, $k = 3$ esetén

$$a_o = \sqrt{\frac{\pi^2}{3} - \frac{49}{36}}, \quad b_o = \frac{11}{6} - \ln(4), \quad a_1 \approx 1.4576, \quad b_1 \approx 1.0643 \quad (5.4.17)$$



5.4.2. ábra. A véges méret korrekció függvénye, $\tilde{P}_1^{\text{var}}(x)$ (5.4.16), gamma mintaméret eloszlás $\langle \phi(y) = \frac{1}{15}(4y)^3 e^{-4y}$ esetén a (2) típusú szten derdizációban. Alakja nagyon hasonló a fix N esetén kapott korrekció függvényéhez ($P_1^{\text{var}}(x)$, (2.6.14)). A véges méret korrekció a $\gamma = \gamma' = 0$ esetre vonatkozik.

A véges méret korrekció függvénye gamma mintaméret eloszlás esetén nagyon hasonló a fix N esetén előálló FSC függvényéhez, amint azt az 5.4.2. ábrán látjuk, a fix N FSC függvény egy jó becslése annak. Amplitúdója, ϵ_{N_0} szintén a fix N esetén felírt módon kapható meg, ahol $N_0 = \langle N \rangle$, az átlagos mintaméret, jelen esetben az általánosított exponenciális öseloszlásnak megfelelően (5.4.11). Vagyis a tény, hogy fix N mintaméret helyett egy skálázó gamma mintaméret eloszlásból húztuk az N értékeket ($\phi(y) = \frac{1}{15}(4y)^3 e^{-4y}$)

megváltoztatta a határeloszlás alakját (5.4.1. ábra), viszont az ehhez a határeloszláshoz tartozó véges méret korrekció közel megegyezik az FTG-eloszláshoz tartozó korrekcióval, amit fix $N = \langle N \rangle$ esetén kapnánk. A véges méret korrekció függvényét kiszámoltuk a tiszta exponenciális őseloszlásra is, ami az FTG osztályban ($\gamma = 0$) kivételes a konvergencia szempontjából, mivel $\gamma' = -1$. A tapasztalat itt is azt mutatta, hogy a gamma eloszlású véletlen N nem változtatta meg lényegesen a korrekció függvényét.

A fenti számolás során egy, a vizsgálni kívánt esetben ideálisabb esetet tekintettünk, ahol a mintaméret eloszlása egy paraméterrel skálázható. A konkrét empiria során viszont a mintaméret eloszlását két-paraméteres gamma eloszlással lehetett jól fittelni (5.3.1)

$$\hat{P}_{N_0}(N) = \frac{1}{N_0 \Gamma(k+1)} \left(\frac{N}{N_0} - b \right)^k e^{-\frac{N}{N_0} + b}, \quad (5.4.18)$$

(itt az $\langle N \rangle$ -gal való skálázás helyett egy formailag egyszerűbb alakot vizsgálunk, mivel $\langle N \rangle$ konstanssal való szorzása a végeredményben csupán a skálázási paramétereket változtatja meg) vagyis a skálázott mintaméret eloszlás

$$\phi(y) = \frac{1}{\Gamma(k+1)} (y-b)^k e^{-y+b}. \quad (5.4.19)$$

Vizsgáljuk meg, miként változtatja meg az extra paraméter a határeloszlást. (5.4.5)-tel analóg módon számolva kapjuk, hogy a határeloszlás

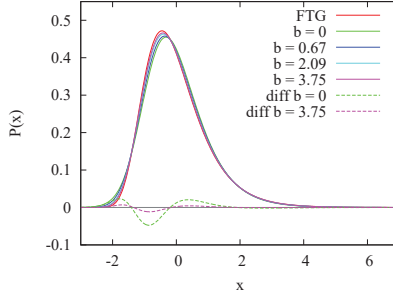
$$\widetilde{M}(x) = \frac{e^{-be^{-x}}}{(1+e^{-x})^{k+1}}, \quad (5.4.20)$$

sűrűségfüggvénye pedig

$$\tilde{P}(x) = \frac{\left[k+1+b(1+e^{-x}) \right] e^{-be^{-x}-x}}{(1+e^{-x})^{k+2}}. \quad (5.4.21)$$

Megjegyezzük még, hogy általános b mellett numerikus úton megvizsgáltuk a véges méret korrekció függvényét, és azt tapasztaltuk, hogy nem változtatja lényegesen annak alakját és amplitúdóját, tehát az FTG-hez tartozó korrekció megfelelő közelítésnek tekinthető, és ezért itt sem térünk ki rá részletesen.

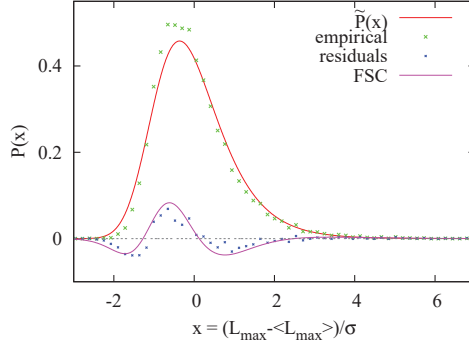
Az alkalmazásunkban érdekes b értékekre, vagyis az átlaggal összekalázott galaxiszám eloszlások illesztési paramétereire (5.3.3.táblázat), mutatja az 5.4.3. ábra a sűrűségfüggvényeket a (2) sztemderizációban. Kis b értékekre a határeloszlás az FTG-től eltér, különbségük az FSC alakjához hasonló. Vegyük észre, hogy $b \rightarrow \infty$ esetén a határeloszlás az FTG-hez



5.4.3. ábra. Extrém határeloszlások két-paraméteres gamma minta a méret eloszlás esetén (5.4.18), az átlaggal összeskálázott galaxisszám eloszlásokra illesztett b értékekkel. A határeloszlás véges b esetén különbözik az FTG-tól, $b \rightarrow \infty$ esetén viszont ahhoz tart, az FTG-tól való eltérése véges méret effektusként jelenik meg (szaggatott görbék).

tart, hiszen (5.4.18) eloszlás alapján a mintaméret várható értéke $\langle N \rangle = (k + 1 + b)N_0$, míg szórása $\sigma_N = \sqrt{k + 1}N_0$, tehát rögzített k -ra relatív szórása $\frac{\sigma_N}{\langle N \rangle} \rightarrow 0$ ha $b \rightarrow \infty$. Az 5.3.3. táblázat adatai szerint a pixelek növelésével, vagyis $N_{side} = 64$ -től 16 felé haladva a gamma eloszlás két illesztési paramétere $N_0 = \frac{1}{a}$ és b is egyre nagyobb, ezzel összhangban a relatív szórás csökken (l. 5.3.1. táblázat). Azzal a feltételezéssel élve, hogy ez egy szisztematikus tendencia, vagyis $N_0 \rightarrow \infty$ -nel $b \rightarrow \infty$ – a mintaméret skálázott eloszlása delta-függvényhez tart – a maximális luminozitások határeloszlása az FTG-eloszlás lesz. Ebből az is következik, hogy a b paraméter is tekinthető egyfajta véges méret paraméterként. Míg a természetes véges méret paraméter, N_0 a tipikus (átlagos) N -t jellemzi, addig b a véges N_0 esetén fennálló $\hat{P}_{N_0}(N)$ eloszlás elkentségére, tehát elsősorban relatív szórására utal.

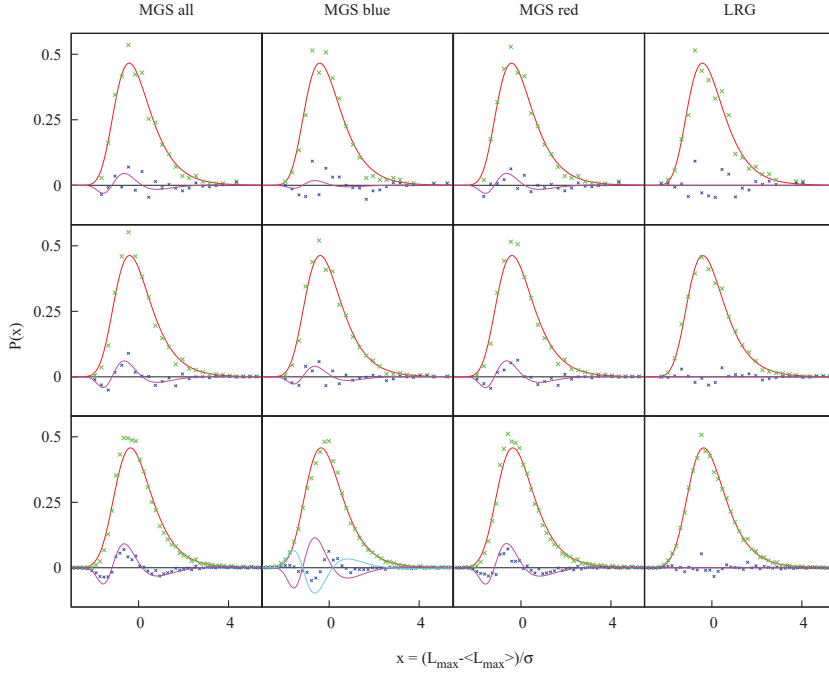
Az empirikus maximális luminozitás eloszlások leírásához a határeloszlás (FTG) mellett véges N_0 esetén tehát két korrekciót kell figyelembe venni. Az egyik a véges N értékek miatt fellépő korrekció, melynek alakját az FTG-eloszláshoz tartozó véges méret korrekció jól leírja, $\epsilon_{N_0} P_1(x) = \epsilon_{N_0} \partial_x \left(-\frac{x^2}{2} e^{-x} - e^{-x} \right)$, ahol N_0 az előforduló N -ek tipikus értéke (átlaga). A másik a mintaméret eloszlás nem pontos skálázásából, jelen esetben az eloszlás relatív szélességéből adódó korrekció, ennek egzakt alakjával itt nem foglalkozunk, de nagyságát becsülhetjük a véges b esetén kapott $\hat{P}(x)$ és az



5.4.4. ábra. Az empirikus maximális luminozítás eloszlás (az ábrán MGSall, $N_{side} = 64$) alakját jól leírja az elméleti random N határeloszlás, $\tilde{P}(x)$, amennyiben az elsőrendű véges méret korrekciót is figyelembe vesszük. Az empirikus hisztogram (zöld) és (5.4.21) határeloszlás (piros) különbségfüggvénye (kék) amplitúdóját és függvényalakját tekintve is jól magyarázható a véges méret korrekcióval.

FTG-eloszlás különbségéből. Mivel az FSC előjele az őseloszlás exponenseitől függ, amplitúdója pedig a tipikus N -tól, bizonyos esetekben a két véges méret korrekció éppen kiolthatja egymást, ezt a tapasztalatot a szimulációk is megerősítik, amiről bővebben a következő alfejezetben lesz szó.

Mivel a b -hez tartozó korrekció alakját analitikusan nem ismerjük, csupán a véges b esetén kiszámolt $\tilde{P}(x)$ határeloszlást, legegyszerűbb, ha az empirikus eloszlásokat ezzel a $\tilde{P}(x)$ határeloszlással vetjük össze. A fenti eredményeknek az empiriával való összevetése alapján azt mondhatjuk, hogy a kapott határeloszlás (5.4.21) a véges méret korrekció figyelembevételével jól magyarázza a megfigyelt maximum luminozítás eloszlásokat. Az 5.4.4. ábrán az MGSall populáció, vagyis a teljes Main Galaxy Sample, cella maximumainak hisztogramját láthatjuk $N_{side} = 64$ felbontás mellett (ezen a felbontáson a legkisebb a statisztikai hiba, és legkisebb az $\langle N \rangle$, emiatt a legjobban látható a véges méret korrekció alakja), a (2) sztenderdizációban, tehát nulla várható értékre és egységnyi szórásra skálázva. A felbontásnak megfelelő elméleti határeloszlás (5.4.21) formula által adott $\tilde{P}(x)$, $b = 0.67$ paraméterrel (5.3.3. táblázat), az ábrán a piros görbe. Az empirikus hisztogram és a határeloszlás



5.4.5. ábra. Empirikus maximális luminozitás hisztogramok (zöld) és a véletlen mintaméret mellett előálló elméleti határeloszlás (piros), $\tilde{P}(x)$, aminek alakját (5.4.21) írja le a (2) sztenderdizációban. A kettő különbsége (kék) jól magyarázható az FTG-hez tartozó véges méret korrekcióval (magenta), amit az átlagos mintamérettel közelítünk. Fentről lefelé $N_{side} = 16, 32, 64$, a zaj mértéke csökken, de nő a véges méret korrekció. Az MGSblue, $N_{side} = 64$ esetében, ahol a legkisebb az átlagos mintaméret, az eloszlás középső tartományára illesztett paraméterek adják az empiriát magyarázó véges méret korrekciót (világos kék). Az LRG esetén nincs elméleti véges méret korrekció az FTG-eloszlás miatt.

különbsége (kék) jól magyarázható az átlagos mintaméretből következő elsőrendű véges méret korrekcióval (magenta), melynek alakja $\epsilon_{\langle N \rangle} \tilde{P}_1(x)$, ahol $\epsilon_{\langle N \rangle}$ -t (5.4.11) határozza meg ($\alpha = -0.81$, $\beta = 0.75$, $\langle N \rangle = 36$), és $\tilde{P}_1(x)$ (5.4.15) által adott.

Az összes vizsgált galaxis populáció és felbontás esetét foglalja össze az 5.4.5. ábra. Ezen az ábrán fittelési paraméter nélkül hasonlítjuk össze az empirikus maximum luminozitás eloszlásokat a véletlen mintaméret esetén előálló elméleti határeloszlással $\tilde{P}(x)$ -el², úgy hogy mindkettőt nulla várható értékűre és egységnyi szórásúra skáláztuk. Az empirikus eloszlás és elméleti határeloszlás különbségét a véges méret korrekció elméleti alakjával vetjük össze. Összességében megállapítható, hogy $N_{side} = 16$ felbontás mellett a hisztogram hibája a véges méret korrekció nagyságrendjébe esik, így az FSC itt nem mérhető. $N_{side} = 32$ és 64 esetében az MGS galaxisaira az empirikus eloszlás és $\tilde{P}(x)$ határeloszlás különbsége a legtöbb esetben mind nagyságrend, mind függvényalak tekintetében jól magyarázható az elsőrendű véges méret korrekcióval (magenta), a statisztikai hiba ellenére is, aminek amplitúdóját az 5.2.1. táblázat illesztési paraméterei határoznak meg (5.4.11) alapján. Egyedül az MGSblue, $N_{side} = 64$ esetében tapasztalunk ellentétes előjelet, ennek okát a kis mintaméretekből találhatjuk (az átlagos mintaméret itt csupán $\langle N \rangle = 13$). A luminozitások várható maximuma $\langle N \rangle = 13$ mellett abba a tartományba esik, ahol a luminozitás függvény és a luminozitás eloszlás szétválik, vagyis az őseloszlás azon tartományába, amelyet az aszimptotikus fit már nem ír le. A luminozitás eloszlás (ami az őseloszlás középső tartományát kellő pontossággal írja le) illesztési paramétereit (5.2.2. táblázat) használva az FSC amplitúdójában, helyes előjelű véges méret korrekciót kapunk (világos kék görbe), ami alátámasztja az empirikus megfigyeléseket.

Kivételes az LRG-k esete, mivel itt az őseloszlás az FTG-eloszlás volt, amelynek nincs véges méret korrekciója, az extrém eloszlás tetszőleges kis N -re egzaktul a határeloszlás. Ez belátható az RG megközelítés segítségével, hiszen eszerint a véges méret viselkedés nem más, mint az RG transzformáció hatása a fixpont közelében egy sajátirány mentén, az FTG-eloszlás pedig a renormálási transzformáció fixpontja, így ebből az őseloszlásból kiindulva a transzformáció nem változtatja az eloszlást, nincs véges méret effektus. Az összefoglaló ábrán látott eredmény is ezt tükrözi, az LRG galaxisaira az empirikus hisztogram és $\tilde{P}(x)$ között csak zajszerű eltérést látunk.

²megj.: a későbbiekben egy hasonló ábrán (5.5.1. ábra) az empirikus eloszlásokat az FTG-eloszlással vetjük össze, és a véletlen mintaméretből adódó különbséget korrekcióként értelmezzük, és a véges méret korrekcióval együtt kezeljük.

5.5. Maximális luminozítás eloszlás meghatározása szimulációval

Bár az extrém eloszlás és véges méret korrekciója analitikusan számolható, csak az $N_0 \rightarrow \infty$ határeloszlást tudjuk meghatározni, és a hozzá tartozó első rendű korrekciót, ami szintén az $N_0 \rightarrow \infty$ limeszben érvényes. Az empirikus görbék jobb leírása érdekében ezért numerikus szimulációkat is végeztünk a megfigyelt N_0 eloszlások figyelembe vételével. A szimulációk során a maximum luminozítás eloszlásfüggvényeket a megfigyelt luminozítás eloszlások mellett nagy statisztikai sokaságra generáltuk, hogy a statisztikai zajt kiküszöbölve az empirikus eloszlások legjobb közelítését kapjuk. Mivel ismert őseloszlásból és mintaméret eloszlásból indultunk ki, a szimuláció az analitikus eredmények ellenőrzésére is alkalmas.

5.5.1. A szimuláció algoritmusa

A szimuláció elvi vázlata a következő. Adott paramétereket tartalmazó $\hat{P}_{N_0}(N)$ mintaméret eloszlásból húzunk egy N értéket, ez lesz a mintaméret. Ezután adott $P(L)$ őseloszlásból húzunk N darab luminozítás értéket $-L_1, L_2, \dots, L_N$ – és kiválasztjuk a maximumot, $L_{\max} = \max\{L_1, L_2, \dots, L_N\}$. Ezeket a lépéseket n -szer ismételve, a kapott L_{\max} értékekből hisztogramot készítünk, amit normálunk, majd nulla várható értékre és egységnyi szórássá skálázunk.

Mintaméret eloszlásként az empirikus galaxisszám eloszlásokra illesztett gamma eloszlásokat használtuk, az 5.3.3. táblázat paramétereivel. Mivel N olyan véletlen változó ami csak egész értékeket vehet fel, az egyszerűbb és gyorsabb szimuláció érdekében a következő diszkrét eloszlást használtuk: az $N = 1, 2, \dots, N_{\max}$ mintaméret értékek előfordulási száma $[n\hat{P}_{N_0}(N)]$ egésze része volt, így biztosítva a kívánt, $\hat{P}_{N_0}(N)$ eloszlású véletlen mintaméretet, ezzel együtt az összes szimuláció száma $\approx n$.

Az őseloszlás a luminozítás függvény aszimptotikájára fittelt általánosított Schechter eloszlás volt. Mivel az extrém értékek úgyis a nagy argumentumú részből kerülnek ki, nem szükséges, hogy a használt őseloszlás pontosan megegyezzen az empirikussal, elég ha az aszimptotikája ugyanaz. Egy adott eloszlású változót egyenletes eloszlású véletlen értékből tudunk generálni, $P_{\text{parent}}(y)dy = P_{\text{uniform}}(x)dx$ megtartásával, vagyis integrálva mindkét oldalt $-M_{\text{parent}}(y) = x - \text{látható, hogy a kívánt eloszlású változót az}$
 $y(x) = M_{\text{parent}}^{-1}(x)$ összefüggésből kaphatjuk meg, ahol x egyenletes eloszlású a $(0, 1)$ intervallumon. Adott őseloszlásból tehát egy megfelelő bin mérettel

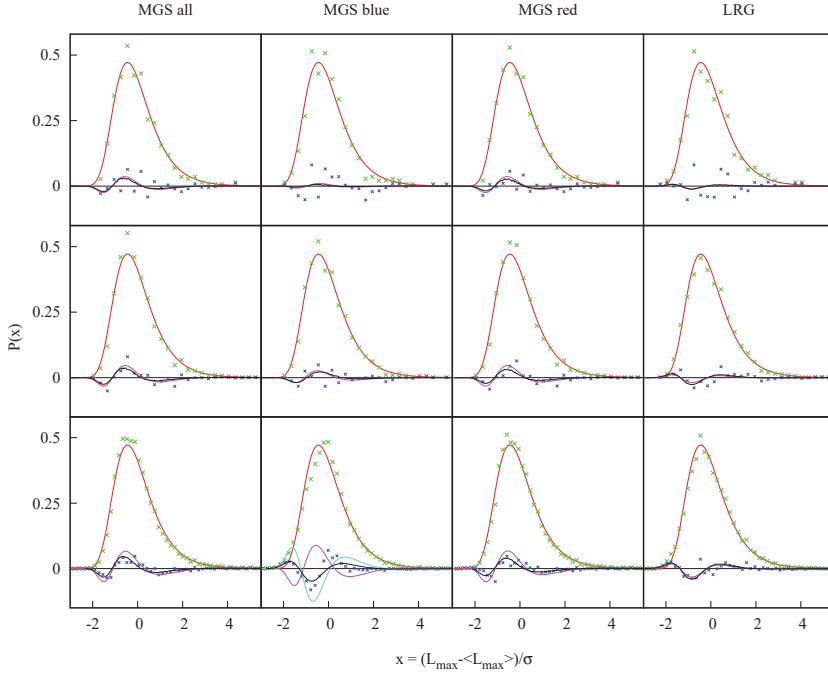
létrehozunk egy diszkrét $M_{\text{parent}}(y)$ eloszlást, melynek inverzéből leolvasható a generált x_1, x_2, \dots, x_N értékekhez tartozó y_1, y_2, \dots, y_N . Az extrém statisztikához azonban csak az y értékek maximumaira van szükségünk, és mivel $M_{\text{parent}}(y)$ – eloszlás lévén – monoton növekvő függvény, elegendő az x_{max} -nak megfelelő y_{max} meghatározása, ezzel tovább csökkentve a futás időigényét.

A szimulációt MATLAB script-ben írtuk meg, random generátorként annak beépített függvényét használtuk. Az eredményül kapott maximum luminozítás hisztogramok statisztikai tulajdonságait illetően, a binek száma ≈ 200 , a hisztogram elkészítéséhez használt maximum értékek, vagyis a futások száma $n \approx 10$ millió volt.

5.5.2. Maximális luminozítás eloszlás szimulálása

A nagyszámú szimulációnak köszönhetően sima eloszlásfüggvényeket tudunk generálni a megfigyelt populációknak és felbontásoknak megfelelő paraméterek mellett az empirikus eredmények modellezésére. A szimulációk során mintaméret eloszlásként az 5.3. fejezetben az egyes galaxisszám eloszlásokra fittelt (5.3.1) gamma eloszlásokat használtuk az 5.3.3. táblázat paramétereivel, őseloszlásként az 5.2. fejezetben fittelt (5.2.3) általánosított Schechter aszimptotikát az 5.2.1. táblázat α és β paramétereivel, illetve az 5.2.2. táblázat – középső tartományra fittelt – paramétereit az MGSblue, $N_{\text{side}} = 64$ esetében.

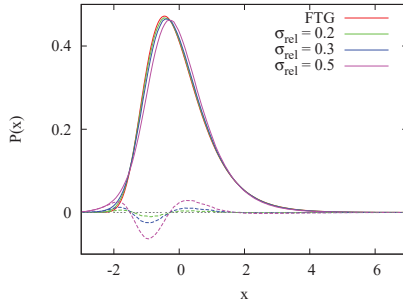
A szimulált maximum luminozítás eloszlások jó egyezést mutattak az empirikusan megfigyelt eloszlásokkal, amint az az 5.5.1. ábrán is látszik, ahol azok FTG-től való eltérését vetettük össze. Az előző, 5.4.5. ábrával ellentétben, ahol a véletlen mintaméret mellett kapott határeloszlással hasonlítottuk össze az empirikus eloszlásokat, itt (a korábbi kutatásoknak megfelelően) az FTG-eloszlással (piros) vetjük össze az empiriát (zöld). A véletlen N hatását korrekcióként értelmezzük, és ezt a korrekciót a véletlen mintaméretű határeloszlás és az FTG különbségével közelítjük. A random N korrekciót a véges méret korrekcióval együtt kezeljük, a két korrekció összegét mutatják a magenta görbék. A szimuláció (fekete) jól modellezi a megfigyelt maximum luminozításokat (kék), ez elsősorban a kisebb zajjal jellemzett $N_{\text{side}} = 32$ és 64 felbontások mellett látszik. Ezenkívül a szimulációk a határeloszláshoz tartozó elméleti korrekciót is alátámasztják (magenta), ami a véges méret korrekció és a véletlen mintaméretből adódó korrekció összege, mitöbb jól látható az elméleti korrekciós függvényhez tartó konvergencia is N_0 értékének növelésével (az ábrát tekintve letről felfelé az egyes populációkra). Az MGSblue, $N_{\text{side}} = 64$ esetében az elméleti korrekcióhoz (világos kék) hasonlóan a szimulációkban is az őseloszlás középső tartományának illesztési



5.5.1. ábra. Szimulált és empirikus maximum luminositás eloszlások összevetése, az FTG-től vett különbségeket hasonlítva össze. Fentről lefelé $N_{side} = 16, 32, 64$. Az előző, 5.4.5. ábrához hasonlóan az empirikus maximális luminositás hisztogramokat mutatják a zöld pontok, viszont attól eltérően a piros görbe itt az FTG eloszlás (mindkét eloszlás a (2) sztenderdizációban van). A két különbsége (kék) egybeesik az FTG-hez tartozó korrekció függvényével (magenta, világos kék), ami itt a véges méret korrekció és a random N korrekció összege. Az MGSblue, $N_{side} = 64$ esetben ismét az őseloszlás középső tartományára illesztett paraméterek adják az empiriával megegyező korrekciót (világos kék). A szimulációk (fekete) jól modellezzik az empirikus eloszlás FTG-től való eltérését (kék), és az elméleti korrekcióval is összhangban vannak. Sőt, megfigyelhető az $N_0 \rightarrow \infty$ konvergencia is (N_0 lentől felfelé nő) az elméleti korrekcióhoz. Az LRG esetén, mivel véges méret korrekció nincs, csak a random N korrekció jelenik meg.

paramétereit használtuk (5.2.2. táblázat), ezzel az empirikus korrekciót jól modellező szimulált eredményeket kaptunk.

Az előző fejezetben kiszámolt elméleti határeloszlás ellenőrzése is rendkívül egyszerű numerikus úton, mivel amennyiben az FTG-eloszlásból indulunk ki, mint őseloszlásból, akkor nincs véges méret korrekció, így véges N_0 értékkel szimulálva is a határeloszlást kapjuk. Az ellenőrzés megmutatta, hogy mind az MGS, mind az LRG galaxisszámaira valóban jó modell a gamma eloszlás, mert a szimulált határeloszlás alakja megegyezik (5.4.21) formuláéval (pontosabban a szimulált hisztogram és az elméleti eloszlás közötti különbség kisebb, mint 1%).



5.5.2. ábra. Extrém határeloszlás adott relatív szórású mintaméret eloszlás esetén. A 4.2.2. fejezet szerint, ha a relatív szórás nullához tart, a határeloszlás a fix N -nek megfelelő lesz. Itt $\gamma = 0$, tehát $\sigma_{\text{rel}} \rightarrow 0$ esetén $\tilde{P}(x) \rightarrow P_{\text{FTG}}(x)$. A szaggatott görbék a határeloszlás és az FTG-eloszlás különbségét mutatják.

5.5.3. Nem skálázó mintaméret esetének vizsgálata szimulációkkal

Szimulációk segítségével lehetőségünk van megvizsgálni a mintaméret eloszlás relatív szórásának hatását az extrém határeloszlásra. A 4.2.2. fejezetben leírtak szerint amennyiben a mintaméret eloszlás, $\hat{P}_{N_0}(N)$, nem skálázható, egy N_0 tipikus értékkel skálázva delta függvényhez tart $N_0 \rightarrow \infty$ határesetben, ez egyben azt is jelenti hogy a relatív szórása $\frac{\sigma_N}{\langle N \rangle} \rightarrow 0$, akkor

az extrém határeloszlás, $\tilde{P}(x)$ változatlanul az őseloszlásnak megfelelő $P(x)$ lesz, $\gamma = 0$ esetén az FTG-eloszlás.

Ennek illusztrálására szimulációkat hajtottunk végre adott relatív szórási mintaméret eloszlással. A modell mintaméret eloszlás egy Gauss-eloszlás volt, $\langle N \rangle = 10$, és 0.2, 0.3 illetve 0.5 relatív szórással (a galaxiszám eloszlások tipikus relatív szórássai), az őseloszlás pedig a véges méret effektus elkerülése érdekében az FTG-eloszlás. A kapott hisztogramok így az adott relatív szórással esetén fellépő $\tilde{P}(x)$ határeloszlások, az 5.5.2. ábrán láthatjuk hogyan tart $\tilde{P}(x)$ az FTG-eloszláshoz, ha $\frac{\sigma_N}{\langle N \rangle} \rightarrow 0$. Szemmel látható a különbségfüggvények hasonlósága az 5.4.3. ábrával, ahol a gamma mintaméret eloszlás esetén megjelenő extrém határeloszlásokat mutattuk be. Ez a megfigyelés numerikus úton támasztja alá a 4.2.3. fejezetben vázolt feltevést, miszerint létezik egy univerzális random N korrekció függvény, amely nem függ a mintaméret eloszlás konkrét alakjától, csupán amplitúdója függ annak relatív szórásától.

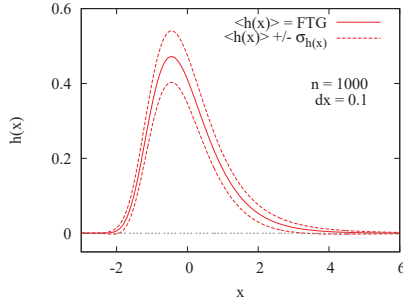
5.6. Hisztogramok hibájának becslése

A maximum luminozítás eloszlások összesítő ábráján (5.4.5. ábra) azt láthattuk, hogy a statisztikai zaj lényeges szerepet játszik a nagy adatbázis ellenére is, olyannyira hogy az $N_{side} = 16$ felbontás esetén teljesen el is fedti a korrekció függvény alakját. Egyszerű megfontolásokkal megbecsülhetjük, hogy egy adott véletlen változó mérése során mekkora a felvett hisztogram hibája.

Tekintsünk egy x változót, melynek eloszlása $P(x)$. Empirikus úton szeretnénk meghatározni ezt az eloszlást, ehhez n számú mérést végzünk, más szavakkal $P(x)$ eloszlásból húzunk n -szer. Egy húzás eredménye x érték dx környezetébe kerül $P(x)dx$ valószínűséggel. Az n darab húzásból hisztogramot készítettünk, dx binnmérettel, ezt nevezzük $h(x)$ -nek. Jelölje (x, dx) az x érték körüli dx méretű bint. Ezt a bint tekintve, egy húzás kimenetele kétféle lehet: $P(x)dx$ valószínűséggel esik a kiválasztott binbe, és $1 - P(x)dx$ valószínűséggel azon kívül. Jelölje $k(x)$ az (x, dx) -be eső mérések számát. Az x mennyiség $P(x)$ eloszlását az így kapott hisztogrammal közelítjük

$$h(x) = \frac{k(x)}{ndx} \quad (5.6.1)$$

normálás után. A leírásból látható, hogy $k(x)$ értéke n húzás után binomiális eloszlást fog követni, ami nagy n értékekre közelíthető Poisson-eloszlással (az ndx mennyiség állandó értéken tartható az $n \rightarrow \infty$ határátmenetben), tehát



5.6.1. ábra. Egy FTG-eloszlású változó véges n mérés által felvett hisztogramja ($h(x)$) a szaggatott görbékkel becsült hibahatárok közé esik, ha a mérések száma $n = 1000$, és a binméret $dx = 0.1$. A hibahatárokat az adott binbe eső érték egyszeres szórásával becsültük ($\langle h(x) \rangle \pm \sigma_{h(x)}$).

átlaga és a szórása

$$\langle k(x) \rangle = nP(x)dx \quad \sigma_{k(x)}^2 = nP(x)dx, \quad (5.6.2)$$

vagyis $h(x)$ hisztogram átlaga és szórása

$$\langle h(x) \rangle = p(x) \quad \sigma_{h(x)} = \sqrt{\frac{P(x)}{ndx}}, \quad (5.6.3)$$

mivel

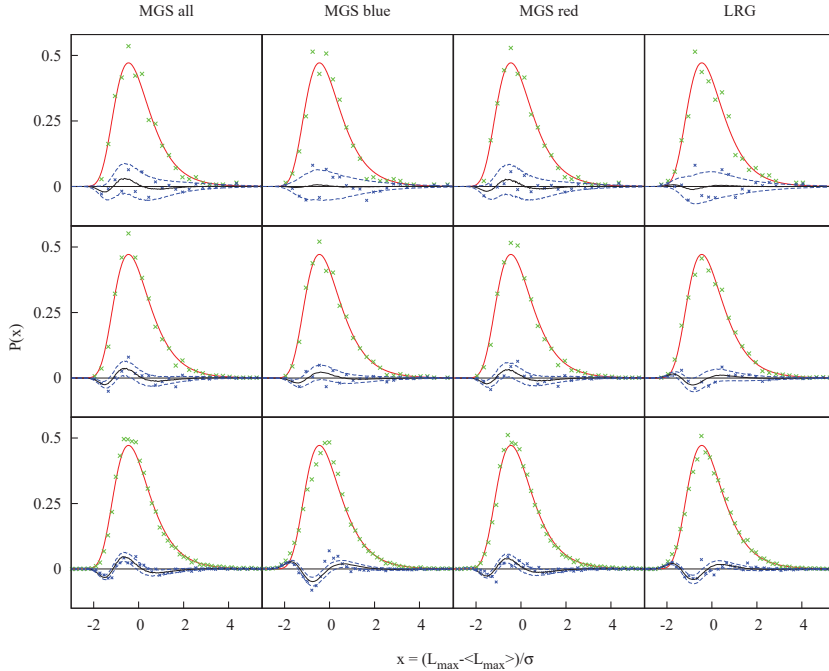
$$\sigma_{h(x)}^2 = \frac{\sigma_{k(x)}^2}{(ndx)^2} = \frac{P(x)}{ndx}. \quad (5.6.4)$$

A fentiek illusztrálására tekintsünk egy FTG-eloszlású változót. $n = 1000$ mérés alapján, $dx = 0.1$ binezéssel felvett hisztogramjának hibáját az 5.6.1. ábrán láthatjuk, ahol az egy szóráson belüli ($\pm\sigma$) hibahatárt a szaggatott görbék jelölik.

A megfigyelt maximum luminozítás eloszlások hibáját ezek alapján becsültük az egyszeres szóráshatárokkal, ahol a hisztogramokhoz felhasznált adatok számát és a binméreteket az 5.6.1. táblázat tartalmazza. A szimulált véges méretű eloszlások korrekciófüggvényei és azok hibahatárai az 5.6.2. ábrán láthatók. Megfigyelhetjük, hogy az empirikus különbségfüggvények nagyrészt az egyszeres szóráson belül találhatók.

	n	dx	ndx
$N_{side} = 64$	8432	0.18	1538.2
$N_{side} = 32$	2030	0.3	609
$N_{side} = 16$	473	0.3	141.9

5.6.1. táblázat. Az empirikus maximum luminozitás hisztogramok paraméterei a három különböző N_{side} felbontásra: n – a hisztogram felvételéhez felhasznált maximum luminozitás értékek száma, és dx – a hisztogram bin mérete. (Mivel az egyes galaxis populációkra n értéke közel azonos, itt ezek átlaga szerepel.)



5.6.2. ábra. Az empirikus maximum luminozitás hisztogramok becsült hibája. A határeloszlástól (FTG) való szisztematikus eltérést a véges és véletlen mintaméret okozza, ezt a szimulált korrekcióval közelítjük (fekete). A véges adatmennyiségből származó hiba egy szórásnyi határait a kék szaggatott vonalak jelölik. Fentről lefelé $N_{side} = 16, 32, 64$.

5.7. Gap statisztika és a Tremaine-Richstone arányszámok

A legnagyobb érték statisztikáján túl az asztrofizika területén is nagy érdeklődésre tart számot a sorrend statisztika. Az első legnagyobb érték mellett a második legnagyobb értéket, illetve az első-második érték közötti gap-et és két – a gap-et jellemző arányszámot, az ún. Tremaine-Richstone hányadosokat vizsgálták a legtöbbet.

A gap és a Tremaine-Richstone hányadosok iránti érdeklődést egy mai napig vitatott kérdés motiválja, miszerint a klaszterek legfényesebb galaxisa (a BCG - Brightest Cluster Galaxy) egyszerűen a luminozítás függvény statisztikai maximuma, vagy különleges, olyan értelemben, hogy más eloszlásból származik, mint a többi klaszter galaxis. A legtöbb megfigyelés során tapasztalt gap értéke lényegesen nagyobb volt, mint amit a luminozítás függvény statisztikája alapján várhatnánk [64, 65, 66]. A BCG-k többi klaszter galaxistól eltérő luminozítás eloszlását azok eltérő (galaxis kannibalizmussal történő) evolúciója indokolhatja [67]. A BCG-knél egy jóval tágabb halmaz, az LRG-k vizsgálata során is felmerült a magnitúdó gap kérdése, ahol szintén a vártnál nagyobb átlagos értéket találtak [68]. A legújabb eredmények ezzel szemben azt mutatják, hogy a körületekintően alkalmazott extrém érték statisztika szerint a BCG-k egyszerűen statisztikai maximumok [69, 70].

Az asztrofizikai irodalomban elsősorban magnitúdó értékekben szokás definiálni az első-második közötti gap-et, amire (5.2.1) szerint tudunk a luminozításról áttérni. A vizsgált mennyiség, mivel a magnitúdó definíciója szerint negatív,

$$\Delta_M = M_2 - M_1 = 2.5 \log_{10} (L_1/L_2), \quad (5.7.1)$$

illetve elsősorban ennek várható értéke $\langle \Delta_M \rangle$. A Tremaine-Richstone hányadosok ennek a várható értéknek a változók szórásához való viszonyát adják meg,

$$T_1 = \frac{\sigma_1}{\langle \Delta_M \rangle}, \quad T_2 = \frac{\sigma_M}{\langle \Delta_M \rangle}, \quad (5.7.2)$$

ahol σ_1 az M_1 legkisebb magnitúdó szórása, és σ_M a Δ_M gap szórása. A Scott modellel [71] leírható klaszter galaxisokra az empirikus hányadosok felső határral rendelkeznek, általános ösloszlás (luminozítás függvény) esetén [65]

$$T_1 \geq 1, \quad T_2 \geq 0.82, \quad (5.7.3)$$

ha az ösloszlás a Schechter-profil által leírt eloszlás [66]

$$T_1 \geq 1.16, \quad T_2 \geq 0.88. \quad (5.7.4)$$

A megfigyelések alapján azonban a BCG-k magnitúdó értékeiből számolt TR hányadosok a gyengébb (5.7.3) egyenlőtlenséget sem teljesítették.

A sorrend statisztika elméletének eredményei alapján a hányadosok $N \rightarrow \infty$ határértéke analitikusan meghatározható, eszerint iid változókra az FTG vonzási tartományában [72]

$$T_1 = \frac{\pi}{\sqrt{6}} = 1.28, \quad T_2 = 1. \quad (5.7.5)$$

Bár ezek a határértékek teljesítik az egyenlőtlenségeket, az elméleti meggondolás azonban a valósághoz képest egy jóval ideálisabb esetet feltételez: iid változókat, vagy legfeljebb gyengén korreláltakat, az N mintaméret fix, nem változik klaszterről klaszterre, és az eredmények csak $N \rightarrow \infty$ határesetben érvényesek, véges méretű minta mellett a megfigyelt értékek ettől különböznek a lassú konvergencia miatt. Ezeket a hatásokat a legújabb kutatásokban minimalizálva az egyenlőtlenségeket nem illetve kevésbé sértő eredményre jutottak, a korreláció hatását a luminozitás adatok összekeverésével csökkentve [66], a változó mintaméret hatását pedig N fixálásával kiküszöbölve [70]. A véges méret miatt fellépő korrekciót azonban eddig nem vették figyelembe, ezért itt megvizsgáljuk elméleti megfontolások és szimulációk segítségével, hogy milyen mértékben befolyásolja a véges N a gap értékét.

Ahhoz, hogy a véges méret korrekciót meghatározzuk, a magnitúdó őseloszlásból kell kiindulnunk. Schechter-profil (5.2.2) feltételezve (ami a luminozitás eloszlás jobboldali aszimptotikáját jól írja le), (5.2.1) változó cserével felírjuk a magnitúdó eloszlás baloldali aszimptotikáját. (5.2.1) szerint

$$L = L_{\odot} \exp\left(\frac{M_{\odot} - M}{d}\right), \quad (5.7.6)$$

ahol $d = 2.5/\ln 10$. A legkisebb magnitúdó érték keresése helyett egy előjelcserével $M \rightarrow -M$, továbbra is a változó maximumát keressük. A magnitúdó eloszlás

$$\Phi(M)dM = \frac{\Phi_*}{d} \left(c \exp(M/d)\right)^{\alpha+1} \exp\left(-c \exp(M/d)\right) dM, \quad (5.7.7)$$

ahol $c = L_{\odot} \exp(M_{\odot}/d)/L_*$. Ebből az integrált őseloszlás függvény közelítése nagy M értékek esetén

$$\mu(M) = 1 - \frac{\Phi_*}{d} \left(c \exp(M/d)\right)^{\alpha} \exp\left(-c \exp(M/d)\right), \quad (5.7.8)$$

aminek segítségével a 2.4. fejezetben leírtak alapján meghatározzuk a véges méret korrekció amplitúdóját.

$$g(z) = -\ln\left(-\ln \mu(z)\right) = -\ln\left(\Phi_* c^{\alpha}/d\right) - \alpha z/d + c \exp(z/d), \quad (5.7.9)$$

ebből a $z \rightarrow \infty$ limitben az átskalázás paraméterei

$$b_N = g^{-1}(\ln N) \approx d \ln(\ln N/c), \quad (5.7.10)$$

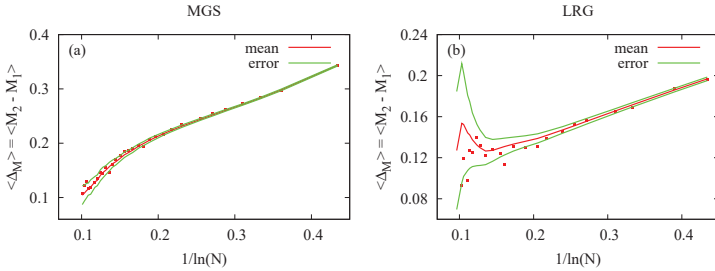
$$a_N = \frac{1}{g'(b_N)} \approx \frac{d}{c} \exp(-b_N/d) \approx \frac{d}{\ln N}, \quad (5.7.11)$$

és az extrém határeloszlást és a véges méret korrekciót meghatározó paraméterek

$$\gamma_N = \frac{da_N}{db_N} = -\frac{1}{c} \exp(-b_N/d) \approx -\frac{1}{\ln N} \rightarrow \gamma = 0 \quad (5.7.12)$$

$$\epsilon_N = \gamma_N - \gamma = -\frac{1}{\ln N}, \quad (5.7.13)$$

vagyis a konvergencia logaritmikus, tehát $\gamma' = 0$.



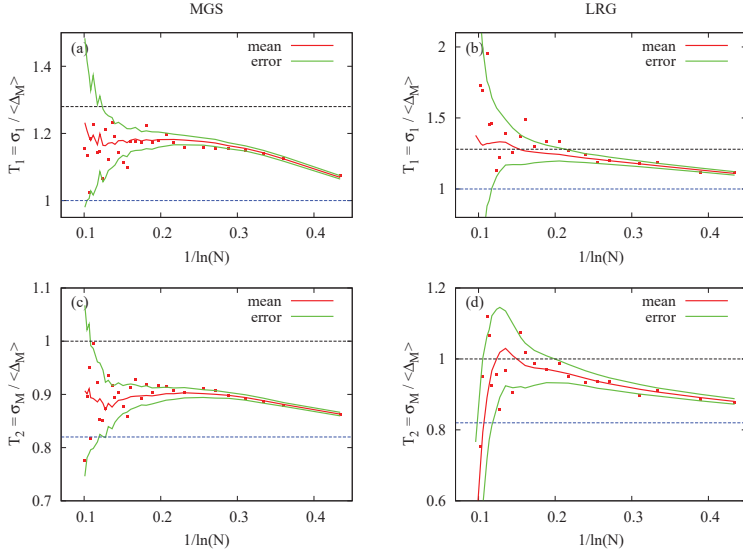
5.7.1. ábra. Magnitúdó gap (az első és második érték távolsága) N mintaméret függése az SDSS galaxisai alapján, az MGS és LRG populációra a mintaméret, a cellákra átlagolva (piros pontok). Az adatokat random permutálva, $n = 1000$ értékből számolt átlagot mutatja a folytonos piros görbe, a zöld az egyszerűes szóráshatárt. A középső N tartományban jól látszik a $\langle \Delta_M \rangle \propto 1/\ln N$ összefüggés.

Beláttuk, hogy a magnitúdó szélsőértékének eloszlása az FTG osztályba tartozik, amire a 3.2. fejezet alapján ismerjük az első-második gap határértékét és véges méret korrekcióját is. A gap határértéke $\Delta_M = 1$ az (1) sztenderdiszidációban, erre a $z = a_N x + b_N$ transzformációval tértünk át, ahol z volt az eredeti változó. Vagyis a gap értéke a magnitúdók eredeti skáláján

$$\Delta_M = a_N = \frac{d}{\ln N} \quad (5.7.14)$$

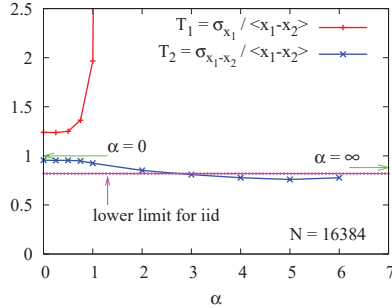
az N mintaméret növelésével logaritmikusan csökken. Ez az eredmény rámutat arra is, hogy nincs értelme a magnitúdó-gap abszolút értékéről beszélni változó elemszámú klaszterek mellett, csak egy átskálázott arányszám lehet

egzakt, összehasonlítható eredmény. A TR hányadosok ennek eleget tesznek, mivel σ_1 és σ_M szórások is $1/\ln N$ -nel divergálnak, ezért T_1 és T_2 határértéke N -tól független mennyiség.



5.7.2. ábra. Tremaine-Richstone hányadosok (a,b) $T_1 = \sigma_1 / \langle \Delta_M \rangle$ és (c,d) $T_2 = \sigma_M / \langle \Delta_M \rangle$ mintaméret (N) függése, az SDSS galaxisai alapján, az MGS és LRG populációra. Mindkét populációra teljesül a $T_1 \geq 1$ illetve $T_2 \geq 0.82$ feltétel, és különösen az LRG-nél, a középső tartományban megfigyelhető az $1/\ln N$ -nel arányos konvergencia a határértékhez ($T_1 \rightarrow 1.28$ illetve $T_2 \rightarrow 1$).

A 3.2. fejezetben beláttuk, hogy a gap véges méret korrekciója is az őseloszlás által meghatározott ϵ_N -nel arányos, a szórással skálázott alakját (3.2.32) adja meg, ami megegyezik T_1 reciprokával. Mivel az elmélet szerint a T_1 hányados logaritmikusan tart az elméleti határértékhez, és a klasztereket alkotó galaxisok száma általában nem túl nagy, a klaszterekben megfigyelhető mennyiség jóval kisebb lehet a határértéknél, vagyis az egyenlőtlenségnek nem feltétlenül kell teljesülnie egy véges méretű mintában. (3.2.32) elméleti véges méret korrekcióval kalkulálva $N \leq 27$ esetén sérül a $T_1 \geq 1$ egyenlőtlenség.



5.7.3. ábra. Tremaine-Richstone hányadosok, T_1 és T_2 korrelált változókra, $1/f^\alpha$ típusú zajokban vizsgálva. A korreláció növelésével (α növelésével) T_2 értéke csökken, míg T_1 nő, és $\alpha = 1$ -nél divergál. A szimulációban $N = 16384$ volt a minta mérete.

A véges méret hatását szimulációkkal is megvizsgáltuk az SDSS MGS és LRG adatai alapján. A korreláció mértékének csökkentése érdekében a magnitúdó adatokat összekevertük, majd egyenlő, N méretű cellákra osztottuk. A cellákra átlagoltuk Δ_M értékeit, és N függvényében vizsgáltuk a konvergenciát, a vizsgált mintaméretnek $N = 10$ -tól $N \approx 20000$ -ig terjedtek. Az 5.7.1. ábrán, bár az egyenes meredeksége nem egyezik az elméleti értékkel, az N középső tartományában világosan látszik a $\langle \Delta_M \rangle \propto 1/\ln N$ összefüggés. A TR hányadosokat vizsgálva az összekevert mintákban azt láthatjuk, hogy az MGS és az LRG galaxisaira megfigyelt értékek az elméleti határértékeknél kisebbek (5.7.2. ábra), de minden vizsgált N mellett teljesítik az (5.7.3) egyenlőtlenségeket. A határértékekhez való konvergencia látható, de egyéb korrekciók illetve a lehetséges korreláció hatása miatt T_1 értéke az egzakt alakkal nem teljesen magyarázható. A határértékek és az FSC alakja iid esetre vonatkozik és bár az adatokat random permutációval összekevertük, lehet köztük korreláció, ami miatt mind a határérték mind a konvergencia sebessége az elmélettől különböző lehet.

A TR hányadosokkal foglalkozó irodalommal ellentmondó eredményünk egyrészt a véges méret effektusokkal, másrészt a korreláció hatásával magyarázható. Az összekevert adatok szimulációja és az elméleti megfontolások alapján azt mondhatjuk, hogy a véges méret effektus a Tremaine-Richstone hányadosok értékét csökkenti. Minél kisebb N , vagyis a klaszter mérete, an-

nál távolabb vagyunk a határértékektől. Bár a szimulációink azt mutatták, hogy még $N = 10$ esetén is teljesülnek az egyenlőtlenségek, az egzakt eredmények szerint azonban megfelelően kis N mellett a véges méret az egyenlőtlenségek sérülését okozhatja. Galaxis klaszterekben ezenkívül a magnitúdók korrelált változók, ezért az iid elmélet eredményeivel nem feltétlenül magyarázhatók. Itt klaszter és field galaxisokat egyaránt vizsgáltunk, a korreláció minimalizálása érdekében összekeverve.

A korreláció hatásának tesztelésére $1/f^\alpha$ típusú zajokban megvizsgáltuk T_1 és T_2 értékét α függvényében. Az α paraméter növelésével a korreláció mértékét növeltük, ahol $\alpha = 0$ az iid, $0 < \alpha < 1$ a gyenge korreláció, míg $1 \leq \alpha \leq \infty$ az erős korrelációnak felel meg (részletesebben l. 3.1. fejezet), és meghatároztuk T_1 és T_2 átlagát. A szimulációk szerint a korreláció mértékének növelésével T_2 hányados értéke csökkent, ellenben T_1 értéke nőtt és $\alpha = 1$ -nél divergált (l. 5.7.3. ábra). Bár a Tremaine-Richstone egyenlőtlenségek sérülését a klaszterekben minden jel szerint a galaxis fényességek korrelációja okozza, más típusú korrelációnak kell jelen lennie, mivel az $1/f^\alpha$ zajokat jellemző korreláció nem magyarázza T_1 alacsony értékét.

6. fejezet

Eredmények összefoglalása

Az extrém érték statisztika gyakorlati alkalmazása során felmerülő problémákat vizsgáltuk dolgozatunk első felében, úgymint véges méret, korreláció, sorrend statisztika és változó méret. Az eredményeket az SDSS galaxis fényesség adatainak extrém statisztikai vizsgálatában használtuk fel a dolgozat második felében, ahol a nagyméretű adatbázis ellenére az eloszlások csak a véges és véletlen méretből adódó korrekciók figyelembe vételével voltak értelmezhetők.

Az általunk bevezetett RG megközelítés azt mutatta, hogy az iid univerzalitási osztályok feloszthatók a véges méret konvergencia exponense (γ') szerint, ami meghatározza a vezető rendű korrekció függvényét is. Az FTG ($\gamma = 0$) osztályon belül felírtuk a véges méret korrekció függvényének egzakt alakját $\gamma' = 0$ (logaritmusos FSC) és $\gamma' = -1$ ($1/N$ -nel arányos FSC) esetén a gyakorlati szempontból lényeges sztenderdizációban (nulla várható érték és egységnyi szórás mellett). Mivel $\gamma' = 0$ tipikus ebben az osztályban, ezért az előbbi korrekciófüggvény az őseloszlások széles körére alkalmazható. Néhány gyakran előforduló őseloszlás példáján mutattuk meg a véges méret korrekció amplitúdójának (ϵ_N) kiszámolását, ami – a tiszta exponenciális esettől eltekintve, ahol $1/N$ -nel arányos – a legtöbb esetben logaritmikusan függ az N véges mérettől. Az FTG-hez logaritmikusan konvergáló őseloszlások közül, három különböző funkcionális formájú véges méret skálázást tárgyaltunk részletesen, a Gauss ($\epsilon_N \sim 1/\ln N$), az e^{-x}/x^α típusú ($\epsilon_N \sim 1/\ln^2 N$) illetve a lognormális ($\epsilon_N \sim 1/\sqrt{\ln N}$) őseloszlás esetét.

A korreláció hatását $1/f^\alpha$ típusú zajok sorrend statisztikájában vizsgáltuk. Az eredményeink alapján a rendezett minta elemei között várható k -adik távolság, $d_k = \langle x_k - x_{k+1} \rangle$ skálázása szerint három tartományt különböztethetünk meg a korreláció függvényében. A szimulációk és elméleti megfontolások szerint az iid skálázás, $d_k \sim k^{-1}$ kiterjeszthető a gyenge korreláció

tartományára, ami $0 < \alpha < 1$ -nek felel meg. Az erős korreláció ($\alpha \geq 1$) tartományában $1 < \alpha < 5$ között egy α -tól függő, $d_k \sim k^{(\alpha-3)/2}$ skálázást találtunk, míg $\alpha > 5$ esetén $d_k \sim k$, α -tól független, egy módus által dominált skálázást tapasztaltunk.

Az első és második legnagyobb érték közötti távolságot (d_1) vizsgálva iid változókra láttuk, hogy annak $N \rightarrow \infty$ határértéke függ az univerzalizációs osztályt jellemző γ paramétertől, és ezért a távolság várható értékének vizsgálata az empirikus adatokban segíthet a határeloszlás azonosításában. Az FTG osztályra meghatároztuk d_1 véges méret korrekcióját, és a legnagyobb érték határeloszlása kapcsán megismert konvergenciát tapasztaltuk, azaz Gauss-típusú őseloszlás esetén $1/\ln N$ -nel, tiszta exponenciális esetén $1/N$ -nel arányos elsőrendű korrekciót találtunk. Az ϵ_N amplitúdó bevezetésével egy általánosított formulával megadtuk, az FTG osztály tetszőleges őseloszlására, a legnagyobb érték szórásával skálázott távolság vezető rendű korrekcióját.

Véletlen mintaméret esetében egzakt úton bemutattuk, hogy adott N -eloszlás mellett az őseloszlás aszimptotikája szerinti szokásos univerzális osztályok megmaradnak, és az extrém eloszlás analóg módon három határeloszlás család egyikéhez fog tartani, amit γ paraméterez. Az N -eloszlás skálázási tulajdonságai határozzák meg, hogy ezek a határeloszlások különböznek-e a fix N határeloszlásoktól. Marad az FTG, FTF és FTW határeloszlás, ha az N -eloszlás, egy tipikus N_0 értékével (pl. átlagával) skálázva, az $N_0 \rightarrow \infty$ limitben gyorsan összeszűkül. Ebben az esetben, véges N_0 mellett, a változó mintaméretből adódó különbség kezelhető véges méret korrekcióként. Egy N_0 paraméterrel, legalább határesetben, skálázó N -eloszlás esetén a skála-függvényről ($\phi(y)$) és γ -tól függő extrém határeloszlást kapunk, a $\phi(y)$ momentum generátor függvényével felírva $G[(1 + \gamma x)^{-1/\gamma}]$ formában. A véges méret korrekció függvénye, pontosan skálázó N -eloszlás mellett, felírható a fix N elmélettel analóg módon, a marginális esetben, függvényalakját az RG csoport kapcsán megismert sajátfüggvény és a véletlen N határeloszlás határozza meg, amplitúdóját az N_0 tipikus értékkel értelmezzük.

Az SDSS által mért galaxis fényesség adatok extrém statisztikáját vizsgálva rámutattunk az alkalmazásokban felmerülő problémákra. Két fő csoportot vizsgáltunk, a Main Galaxy Sample (MGS) és Luminous Red Galaxies (LRG) adatait, az MGS-t tovább bontva MGSblue és MGSred alcsoportokra, szín és morfológia szerint. Az MGS galaxis fényességeinek eloszlása, az őseloszlás, az általánosított Schechter eloszlással (általánosított gamma eloszlás) jól illeszthető, így az extrém eloszlás az FTG vonzási tartományába tartozik. Az LRG-kra, amik eleve környezetük legfényesebb galaxisai, ennek megfelelően az FTG bizonyult a legjobb fitnek. A fényességek EVS analí-

zise alátámasztotta az őseloszlások aszimptotikus fitjét. Mivel adott térszög maximumait vizsgáltuk, a mintaméret változó volt. Az adott irányban megfigyelt galaxisok számának eloszlása azonban az átlaggal összekálázható a különböző populációkra. Mivel az N -eloszlás skálafüggvénye keskeny, és a különböző felbontásokat tanulmányozva $\langle N \rangle$ -gal összeszűkül, ezért az elméleti határeloszlás valóban az FTG. A mintaméret eloszlásra jól fittelő gamma eloszlás mellett analitikus úton meghatároztuk az extrém határeloszlást és véges méret korrekciót, és az eredmények figyelembevételével vizsgáltuk az empirikus eloszlásokat. Mivel a tipikus $\langle N \rangle \sim 10 - 100$ volt, a véges méret és a véletlen N korrekció jelentősnek bizonyult. Az elméleti és a szimulált korrekciók jó egyezése az empirikus adatokkal azt mutatta, hogy a galaxis fényességek extrém statisztikája jól értelmezhető az iid elmélet keretein belül, és a korrelációk a pencil beam geometriában gyengéek.

Megvizsgáltuk az első és második legfényesebb galaxis magnitúdó-gap értékét és a Tremaine-Richstone arányszámokat az MGS és LRG adataiban. A megfigyelés, hogy az összekevert mintában az egyenlőtlenségek minden vizsgált N értékre teljesültek, arra enged következtetni, hogy a klaszterekben jelenlévő erős korreláció az egyik oka az egyenlőtlenségek sérülésének. Szimulációkkal és analitikus úton megmutattuk a lassú, mintamérettel logaritmikus konvergenciát a határértékekhez, az eredmények szerint a véges méret korrekció kis elemszámú klaszterekben szintén hozzájárul a TR arányszámok vártnál alacsonyabb értékéhez.

Az extrém érték statisztikát analitikusan vizsgálva azt találtuk, hogy a véges és a véletlen mintaméretből adódó korrekció lényeges járulékot ad a határeloszláshoz. Rámutattunk, hogy az őseloszlás helyes kezelése fontos, mivel az elméleti határeloszlások és korrekciók érzékenyen függenek annak aszimptotikus viselkedésétől. A galaxis fényességek extrém statisztikai analízise során láttuk, hogy a fenti két korrekció valóban szükséges az empirikus eloszlások megértésében. Azt is láttuk viszont, hogy még a jelenleg legnagyobb galaxis adatbázis sem feltétlenül elég a megbízható extrém statisztikához. Az utóbbi évek tudományos fejlődésének megfelelően az adathalmazok mérete és száma exponenciálisan nő, és így az extrém érték statisztika vizsgálatára egyre több területen nyílik lehetőség. Mivel nagy adathalmaz esetén is jelentős korrekciók lépnek fel, az általunk kidolgozott eredmények lényegesen segíteni fogják ezeket az analíziseket.

A. Függelék

Extrém határeloszlás két speciális esetben

Ez a függelék két olyan mintaméret eloszlás esetén megjelenő extrém határeloszlás számolását vázolja fel, amiknek általános őseloszlás melletti tárgyalását már láttuk a 4.1. fejezetben. Itt adott őseloszlás mellett számoljuk ki az extrém határeloszlást, ami egyrészt előzménye az általános tárgyalásnak, másrészt analitikus volta miatt azt alátámasztja. Az általános tárgyaláshoz képest, ahol nem foglalkoztunk a sztenderdizációs paraméterek kiszámolásával, itt megadjuk azokat nullad és elsőrendben a (2) sztenderdizáció szerint.

A.1. Egyenletes mintaméret eloszlás, exponenciális őseloszlás

Egyszerű és triviális eset, mikor a mintaméret eloszlása egyenletes, vagyis $\hat{P}_{N_0}(N) = \frac{1}{N_0}$, ha $0 < N < N_0$. Ekkor (4.1.3) szerint a maximum eloszlása

$$\tilde{P}_{N_0}^{\max}(z) = \frac{1}{N_0} \int_0^{N_0} P_N^{\max}(z) dN, \quad (\text{A.1.1})$$

ahol $P_N^{\max}(z)$ az N elemű mintából húzott maximum eloszlása. Exponenciális őseloszlás, vagyis $\mu(z) = 1 - e^{-z}$, ($z > 0$) esetén (4.1.2) szerint

$$P_N^{\max}(z) = N e^{-z} (1 - e^{-z})^{N-1}. \quad (\text{A.1.2})$$

Tehát a mintaméret eloszlásával súlyozott extrém eloszlás (A.1.1)

$$\begin{aligned}
\tilde{P}_{N_0}^{\max}(z) &= \frac{1}{N_0} \int_0^{N_0} N e^{-z} (1 - e^{-z})^{N-1} dN = \frac{1}{N_0} \frac{e^{-z}}{1 - e^{-z}} \int_0^{N_0} N e^{N \ln(1 - e^{-z})} dN \\
&= \frac{1}{N_0} \frac{e^{-z}}{1 - e^{-z}} \left(\left[\frac{N e^{N \ln(1 - e^{-z})}}{\ln(1 - e^{-z})} \right]_0^{N_0} - \int_0^{N_0} \frac{e^{N \ln(1 - e^{-z})}}{\ln(1 - e^{-z})} dN \right) = \\
&= \frac{1}{N_0} \frac{e^{-z}}{(1 - e^{-z}) \ln(1 - e^{-z})} \left(N_0 e^{N_0 \ln(1 - e^{-z})} - \left[\frac{e^{N \ln(1 - e^{-z})}}{\ln(1 - e^{-z})} \right]_0^{N_0} \right) = \\
&= \frac{1}{N_0} \frac{e^{-z}}{(1 - e^{-z}) \ln^2(1 - e^{-z})} \left((N_0 \ln(1 - e^{-z}) - 1) (1 - e^{-z})^{N_0} + 1 \right). \tag{A.1.3}
\end{aligned}$$

Egy alkalmas lineáris transzformációval egységnyi szórású és nulla várható értékű alakra skálázzuk az eloszlást, vagyis $z = a_{N_0}x + b_{N_0}$, ahol $\overline{x} = 0$ és $\overline{x^2} = 1$ meghatározza az a_{N_0} , b_{N_0} együtthatókat

$$b_{N_0} = \overline{z}, \quad a_{N_0} = \sqrt{z^2 - b_{N_0}^2}, \tag{A.1.4}$$

és keressük az $N_0 \rightarrow \infty$ esetén a

$$\tilde{P}_{N_0}^{\max}(a_{N_0}x + b_{N_0}) \rightarrow \tilde{P}(x) \tag{A.1.5}$$

határeloszlást. Az x új változó definíciója alapján

$$\begin{aligned}
b_{N_0} = \overline{z} &= \int_0^\infty z \tilde{P}_{N_0}^{\max}(z) dz = \\
&= \frac{1}{N_0} \int_0^\infty \frac{z e^{-z}}{(1 - e^{-z}) \ln^2(1 - e^{-z})} \left((N_0 \ln(1 - e^{-z}) - 1) (1 - e^{-z})^{N_0} + 1 \right) dz.
\end{aligned}$$

Mivel ismert, hogy exponenciális öseloszlás esetén a maximum várható értéke logaritmikusan nő a minta méretével, érdemes végrehajtani a $z \rightarrow z + \ln N_0$ változócserét

$$\begin{aligned}
b_{N_0} &= \frac{1}{N_0^2} \int_{-\ln N_0}^\infty \frac{(z + \ln N_0) e^{-z}}{(1 - \frac{e^{-z}}{N_0}) \ln^2(1 - \frac{e^{-z}}{N_0})} \left((N_0 \ln \left(1 - \frac{e^{-z}}{N_0} \right) - 1) \cdot \right. \\
&\quad \left. \cdot \left(1 - \frac{e^{-z}}{N_0} \right)^{N_0} + 1 \right) dz. \tag{A.1.6}
\end{aligned}$$

Következő lépésként végrehajtjuk az $N_0 \rightarrow \infty$ határátmenetet és sorbafejtünk $\frac{1}{N_0}$ szerint, megtartva az elsőrendű tagokat is $\frac{1}{N_0}$ -ban, ami az elsőrendű véges méret korrekciót fogja adni a paraméterekben. Az integrál tagjainak

sorfejtése

$$\frac{1}{\left(1 - \frac{e^{-z}}{N_0}\right)} \approx 1 + \frac{e^{-z}}{N_0} + \mathcal{O}\left(\frac{1}{N_0^2}\right), \quad (\text{A.1.7})$$

$$N_0 \ln \left(1 - \frac{e^{-z}}{N_0}\right) \approx -e^{-z} - \frac{e^{-2z}}{2N_0} + \mathcal{O}\left(\frac{1}{N_0^2}\right), \quad (\text{A.1.8})$$

$$\frac{1}{N_0^2 \ln^2 \left(1 - \frac{e^{-z}}{N_0}\right)} \approx e^{2z} - \frac{e^z}{N_0} + \mathcal{O}\left(\frac{1}{N_0^2}\right), \quad (\text{A.1.9})$$

$$\left(1 - \frac{e^{-z}}{N_0}\right)^{N_0} = e^{N_0 \ln \left(1 - \frac{e^{-z}}{N_0}\right)} \approx e^{-e^{-z}} \left(1 - \frac{e^{-2z}}{2N_0} + \mathcal{O}\left(\frac{1}{N_0^2}\right)\right). \quad (\text{A.1.10})$$

Tehát a várható érték sorbafejtve és tovább egyszerűsítve

$$\begin{aligned} b_{N_0} &= \overline{z} = \int_{-\ln N_0}^{\infty} (z + \ln N_0) e^{-z} \left(1 + \frac{e^{-z}}{N_0}\right) \left(e^{2z} - \frac{e^z}{N_0}\right) \cdot \\ &\quad \cdot \left(\left(-e^{-z} - \frac{e^{-2z}}{2N_0} - 1\right) e^{-e^{-z}} \left(1 - \frac{e^{-2z}}{2N_0}\right) + 1 + \mathcal{O}\left(\frac{1}{N_0^2}\right)\right) dz = \\ &= \int_{-\ln N_0}^{\infty} (z + \ln N_0) e^z \left(e^{-e^{-z}} \left(-1 - e^{-z} + \frac{e^{-3z}}{2N_0}\right) + 1 + \mathcal{O}\left(\frac{1}{N_0^2}\right)\right) dz = \\ &= \int_{-\infty}^{\infty} (z + \ln N_0) \left(e^z - e^{-e^{-z}} - e^{z-e^{-z}} + \frac{e^{-2z-e^{-z}}}{2N_0}\right) dz + \mathcal{O}\left(\frac{1}{N_0^2}\right), \end{aligned} \quad (\text{A.1.11})$$

ahol az integrálás alsó határát kitöltük $-\infty$ -ig, mert a $(-\infty, -\ln N_0)$ tartomány 0-hoz tartó járulékot ad az $N_0 \rightarrow \infty$ határesetben, mivel mindegyik tag legalább $1/N_0$ szerint tart 0 hoz a $-\infty$ -ben. Az integrálás szerint az eltolási paraméter nullad- és elsőrendben

$$b_{N_0,0} = \ln N_0 + \gamma_E - 1, \quad (\text{A.1.12})$$

$$b_{N_0,1} = \frac{\ln N_0}{2N_0} + \frac{\gamma_E - 1}{2N_0}, \quad (\text{A.1.13})$$

ahol $\gamma_E = 0.5772$ az Euler-állandó. A skálázó paraméter meghatározásához a változó szórásából kell kiindulni

$$\begin{aligned} \overline{z^2} &= \int_0^{\infty} z^2 \tilde{P}_{N_0}^{\max}(z) dz = \\ &= \frac{1}{N_0} \int_0^{\infty} \frac{z^2 e^{-z}}{(1 - e^{-z}) \ln^2(1 - e^{-z})} \left(\left(N_0 \ln(1 - e^{-z}) - 1\right)(1 - e^{-z})^{N_0} + 1\right) dz. \end{aligned} \quad (\text{A.1.14})$$

Az előző gondolatmenetet követve, a $z \rightarrow z + \ln N_0$ változócsere végrehajtva és sorbafejtve $\frac{1}{N_0}$ hatványai szerint

$$\begin{aligned} \overline{z^2} &= \int_{-\infty}^{\infty} (z + \ln N_0)^2 \left(e^z - e^{-e^{-z}} - e^{z-e^{-z}} + \frac{e^{-2z-e^{-z}}}{2N_0} \right) dz + \mathcal{O}\left(\frac{1}{N_0^2}\right) = \\ &= \frac{\pi^2}{6} + \gamma_E^2 - 2\gamma_E + 2 + 2\ln N_0(\gamma_E - 1) + \ln^2 N_0 + \\ &+ \frac{1}{2N_0} \left(\frac{\pi^2}{6} + \gamma_E^2 - 2\gamma_E + 2\ln N_0(\gamma_E - 1) + \ln^2 N_0 \right), \end{aligned} \quad (\text{A.1.15})$$

tehát a skálaparaméter $a_{N_0}^2 = \overline{z^2} - b_{N_0}^2$ nullad- és elsőrendben

$$\begin{aligned} a_{N_0}^2 &= \frac{\pi^2}{6} + 1 + \frac{1}{2N_0} \left(\frac{\pi^2}{6} - \gamma_E^2 + 2\gamma_E - 2 - 2\ln N_0(\gamma_E - 1) - \ln^2 N_0 \right), \\ a_{N_0,0} &= \sqrt{\frac{\pi^2}{6} + 1}, \\ a_{N_0,1} &= \frac{1}{4N_0 a_{N_0,0}} \left(\frac{\pi^2}{6} - \gamma_E^2 + 2\gamma_E - 2 - 2\ln N_0(\gamma_E - 1) - \ln^2 N_0 \right). \end{aligned} \quad (\text{A.1.16})$$

$$(\text{A.1.17})$$

Meghatároztuk tehát a $z = a_{N_0}x + b_{N_0}$ lineáris transzformáció paramétereit, amellyel az extrém határeloszlás az $N_0 \rightarrow \infty$ limeszben egy nem degenerált függvény lesz, nulla átlaggal, és egységnyi szórással, nulladrendben

$$z = \sqrt{\frac{\pi^2}{6} + 1} \cdot x + \ln N_0 + \gamma_E - 1. \quad (\text{A.1.18})$$

Az extrém határeloszlást az új változóra áttérve és az $N_0 \rightarrow \infty$ limeszt elvégezve kapjuk, $a_0 = \sqrt{\frac{\pi^2}{6} + 1}$ és $b_0 = \gamma_E - 1$ jelöléssel

$$\begin{aligned} \tilde{P}_{N_0}^{\max}(x) &= a_{N_0} \tilde{P}_{N_0}^{\max}(a_{N_0}x + b_{N_0}) \rightarrow \tilde{P}(x), \\ \tilde{P}(x) &= a_0 \left(e^{a_0x+b_0} - e^{-e^{-(a_0x+b_0)}} - e^{a_0x+b_0-e^{-(a_0x+b_0)}} \right). \end{aligned} \quad (\text{A.1.19})$$

A.2. Exponenciális mintaméret eloszlás, exponenciális őseloszlás

Második esetként vizsgáljuk meg az exponenciálisan lecsengő mintaméret eloszlást, vagyis $\tilde{P}_{N_0}(N) = \frac{1}{N_0} e^{-N/N_0}$. (4.1.3) szerint a maximum eloszlása

$$\tilde{P}_{N_0}^{\max}(z) = \frac{1}{N_0} \int_0^{\infty} e^{-N/N_0} P_N^{\max}(z) dN, \quad (\text{A.2.1})$$

ahol $P_N^{\max}(z)$ -t exponenciális öseloszlás esetén (A.1.2) adja meg, vagyis

$$\begin{aligned} \tilde{P}_{N_0}^{\max}(z) &= \frac{1}{N_0} \int_0^\infty e^{-\frac{N}{N_0}} N e^{-z} (1 - e^{-z})^{N-1} dN = \frac{e^{-z}}{N_0(1 - e^{-z})} \\ &\cdot \int_0^\infty N e^{-\left(\frac{1}{N_0} - \ln(1 - e^{-z})\right)N} dN = \frac{e^{-z}}{N_0(1 - e^{-z}) \left(\frac{1}{N_0} - \ln(1 - e^{-z})\right)^2}. \end{aligned} \quad (\text{A.2.2})$$

Az előző esettel analóg módon, meghatározzuk a lineáris transzformáció (A.1.4) paramétereit, sorbafejtünk $\frac{1}{N_0}$ hatványai szerint, és felírjuk az (A.1.5) határeloszlást az $N_0 \rightarrow \infty$ határesetben.

$$\begin{aligned} b_{N_0} &= \bar{z} = \int_0^\infty \frac{z e^{-z}}{N_0(1 - e^{-z}) \left(\frac{1}{N_0} - \ln(1 - e^{-z})\right)^2} dz = \\ &= \int_{-\ln N_0}^\infty \frac{(z + \ln N_0) e^{-z}}{N_0^2(1 - \frac{e^{-z}}{N_0}) \left(\frac{1}{N_0} - \ln(1 - \frac{e^{-z}}{N_0})\right)^2} dz, \end{aligned} \quad (\text{A.2.3})$$

ahol, mivel az öseloszlás exponenciális, ismét végrehajtottuk a $z \rightarrow z + \ln N_0$ változócserét. Az (A.1.7)-t, valamint az alábbi sorfejtést felhasználva

$$\frac{1}{N_0^2 \left(\frac{1}{N_0} - \ln(1 - \frac{e^{-z}}{N_0})\right)^2} \approx \frac{1}{(1 + e^{-z})^2} \left(1 - \frac{1}{N_0} \frac{e^{-2z}}{(1 + e^{-z})}\right) + \mathcal{O}\left(\frac{1}{N_0^2}\right), \quad (\text{A.2.4})$$

láthatjuk, hogy a b_{N_0} paraméter

$$b_{N_0} = \int_{-\ln N_0}^\infty (z + \ln N_0) \frac{e^{-z}}{(1 + e^{-z})^2} \left(1 + \frac{1}{N_0} \left(e^{-z} - \frac{e^{-2z}}{1 + e^{-z}}\right) + \mathcal{O}\left(\frac{1}{N_0^2}\right)\right) dz. \quad (\text{A.2.5})$$

Az integrálás alsó határát az $N_0 \rightarrow \infty$ limeszben kiterjeszthetjük $-\infty$ -ig, ekkor elsőrendig

$$b_{N_0} = \int_{-\infty}^\infty (z + \ln N_0) \frac{e^{-z}}{(1 + e^{-z})^2} \left(1 + \frac{1}{N_0} \frac{e^{-z}}{(1 + e^{-z})}\right) dz. \quad (\text{A.2.6})$$

Tehát a b_{N_0} paraméter nullad- és elsőrendű tagja

$$b_{N_0,0} = \ln N_0, \quad (\text{A.2.7})$$

$$b_{N_0,1} = \frac{\ln N_0}{2N_0} - \frac{1}{2N_0}. \quad (\text{A.2.8})$$

A változó négyzetének várható értékét ugyanezen lépések végrehajtásával számoljuk

$$\begin{aligned} \overline{z^2} &= \int_0^\infty \frac{z^2 e^{-z}}{N_0(1 - e^{-z}) \left(\frac{1}{N_0} - \ln(1 - e^{-z}) \right)^2} dz = \int_{-\ln N_0}^\infty \frac{(z + \ln N_0)^2 e^{-z}}{N_0^2 \left(1 - \frac{e^{-z}}{N_0}\right)} \\ &\cdot \frac{1}{\left(\frac{1}{N_0} - \ln(1 - \frac{e^{-z}}{N_0}) \right)^2} dz = \int_{-\infty}^\infty (z + \ln N_0)^2 \frac{e^{-z}}{(1 + e^{-z})^2} \left(1 + \frac{1}{N_0} \frac{e^{-z}}{1 + e^{-z}} \right) dz, \\ \overline{z^2} &= \ln^2 N_0 + \frac{\pi^2}{3} + \frac{\ln^2 N_0}{2N_0} - \frac{\ln N_0}{N_0} + \frac{1}{N_0} \frac{\pi^2}{6}, \end{aligned} \quad (\text{A.2.9})$$

tehát a skálaparaméter $a_{N_0}^2 = \overline{z^2} - b_{N_0}^2$ nullad- és elsőrendben

$$a_{N_0}^2 = \frac{\pi^2}{3} - \frac{\ln^2 N_0}{2N_0} + \frac{1}{N_0} \frac{\pi^2}{6}, \quad (\text{A.2.10})$$

$$a_{N_0,0} = \frac{\pi}{\sqrt{3}}, \quad (\text{A.2.11})$$

$$a_{N_0,1} = \frac{\sqrt{3}}{4\pi} \left(-\frac{\ln^2 N_0}{N_0} + \frac{1}{N_0} \frac{\pi^2}{3} \right). \quad (\text{A.2.12})$$

Az extrém határeloszlás tehát az új változóra áttérve– nulladrendben $z = \frac{\pi}{\sqrt{3}}x + \ln N_0$ -, az $N_0 \rightarrow \infty$ limeszben

$$\begin{aligned} \tilde{P}_{N_0}^{\max}(x) &= a_{N_0} \tilde{P}_{N_0}^{\max}(a_{N_0}x + b_{N_0}) \rightarrow \tilde{P}(x), \\ \tilde{P}(x) &= \frac{\pi}{\sqrt{3}} \frac{e^{-\frac{\pi}{\sqrt{3}}x}}{\left(1 + e^{-\frac{\pi}{\sqrt{3}}x}\right)^2} = \frac{\pi}{\sqrt{3}} \frac{1}{\left(2 \cosh\left(\frac{\pi}{2\sqrt{3}}x\right)\right)^2}. \end{aligned} \quad (\text{A.2.13})$$

B. Függelék

Rövidítések

BCG	Brightest Cluster Galaxy
EVS	Extreme Value Statistics
FSC	Finite Size Correction
FTF	Fisher-Tippett-Fréchet eloszlás
FTG	Fisher-Tippett-Gumbel eloszlás
FTW	Fisher-Tippett-Weibull eloszlás
GEV	Generalized Extreme Value Distribution
HEALPix	Hierarchical Equal Area isoLatitude Pixelisation
iid	independent identically distributed
LRG	Luminous Red Galaxies
MGs	Main Galaxy Sample
MRH	Maximum Relative Height
SDSS	Sloan Digital Sky Survey
SDSS DR8	Sloan Digital Sky Survey Data Release 8
SDSS foot print	az égbolt feltérképezett területe az SDSS által
TR	Tremaine-Richstone

Köszönetnyilvánítás

Mindenekelőtt szeretnék köszönetet mondani témavezetőimnek. Rácz Zoltánnak, aki mindvégig bölcs és alapos megfontoltsággal irányította, segítette munkámat, és akitől a közös munka éve alatt rengeteget tanultam, szakmai útmutatásán túl emberi értékek terén is.

Györgyi Gézának, aki bár doktori tanulmányaim kezdetén még nem volt hivatalos témavezetőm, kérdéseimmel mindig fordulhattam hozzá, időt és energiát nem sajnálva segített tisztázni a problémákat.

Köszönet szerzőtársaimnak, Manushehr Taghizadeh-Popp-nak a galaxis adatok feldolgozásáért, Nicholas Moloney-nak a szimulációk terén nyújtott segítségéért, Michel Droz-nak, Szalay Sándornak és Regős Enikőnek az eredményes közös munkáért. Köszönet Tél Tamásnak munkám támogatásáért és az Elméleti Fizika Tanszék munkatársainak a nyugodt körülményekért.

Köszönöm páromnak, Korbely Attilának, és legfőképpen szüleimnek, hogy hittek bennem és mellettem álltak.

Irodalomjegyzék

- [1] E.J. Gumbel, *Statistics of Extremes* (Columbia University Press, New York, 1958).
- [2] R.A. Fisher and L.H.C. Tippett, *Limiting forms of the frequency distribution of the largest or smallest member of a sample*, Procs. Cambridge Philos. Soc. **24**, 180 (1928).
- [3] J. Galambos, *The Asymptotic Theory of Extreme Order Statistics* (R. E. Krieger, Malabar, Florida, 1987).
- [4] L. de Haan, and A. Ferreira, *Extreme Value Theory: An Introduction* (Springer, New York, 2006).
- [5] W. Weibull, *A Statistical Distribution Function of Wide Applicability*, J. Appl. Mech. **18**, 293 (1951).
- [6] P. Embrecht, C. Klüppelberg and T. Mikosch, *Modeling Extremal Events for Insurance and Finance* (Springer, Berlin, 1997).
- [7] B. Gutenberg and C. F. Richter, *Frequency of earthquakes in California*, Bull. Seismol. Soc. Am. **34**, 185 (1944).
- [8] R. W. Katz, M. B. Parlange and P. Naveau, *Statistics of extremes in hydrology*, Adv. Water Resour. **25**, 1287 (2002).
- [9] H. V. Storch and F. W. Zwiers, *Statistical Analysis in Climate Research* (Cambridge University Press, Cambridge, 2002).
- [10] R. W. Katz and B. G. Brown, *Extreme events in a changing climate: Variability is more important than averages*, Climatic Change **21**, 289 (1992).
- [11] R. W. Katz, *Extreme value theory for precipitation: sensitivity analysis for climate change*, Adv. Water Resour. **23**, 133 (1999).
- [12] J.-P. Bouchaud and M. Mézard, *Universality Classes for Extreme Value Statistics*, J. Phys. A **30**, 7997 (1997).

- [13] T. Antal, M. Droz, G. Györgyi, and Z. Rácz, *1/f noise and extreme value statistics*, Phys. Rev. Lett. **87**, 240601 (2001).
- [14] G. Györgyi, P. C. W. Holdsworth, B. Portelli, and Z. Rácz, *Statistics of extremal intensities for Gaussian interfaces*, Phys. Rev. E **68**, 056116 (2003).
- [15] P. Le Doussal and C. Monthus, *Exact solutions for the statistics of extrema of some random 1D landscapes, Application to the equilibrium and the dynamics of the toy model*, Physica A **317**, 140 (2003).
- [16] S. Raychaudhuri, M. Cranston, C. Przybyla, and Y. Shapir, *Maximal height scaling of kinetically growing surfaces*, Phys. Rev. Lett. **87**, 136101 (2001).
- [17] D.-S. Lee, *Distribution of extremes in the fluctuations of two-dimensional equilibrium interfaces*, Phys. Rev. Lett. **95**, 150601 (2005).
- [18] C. J. Bolech and A. Rosso, *Universal Statistics of the Critical Depinning Force of Elastic Systems in Random Media*, Phys. Rev. Lett. **93**, 125701 (2004).
- [19] E. Bertin, *Global fluctuations and Gumbel statistics*, Phys. Rev. Lett. **95**, 170601 (2005).
- [20] H. Guclu and G. Korniss, *Extreme Fluctuations in Small-Worlds with Relational Dynamics*, Phys. Rev. E **69**, 065104(R) (2004).
- [21] P. L. Krapivsky and S. N. Majumdar, *Traveling Waves, Front Selection, and Exact Nontrivial Exponents in a Random Fragmentation Problem*, Phys. Rev. Lett. **85**, 5492 (2000).
- [22] A. Comtet, P. Leboeuf and S. N. Majumdar, *Level Density of a Bose Gas and Extreme Value Statistics*, Phys. Rev. Lett. **98**, 070404 (2007).
- [23] L. de Haan and S. Resnick, *Second order regular variation and rates of convergence in extreme value theory*, Annals of Probability **24**, 97-124 (1996).
- [24] M. I. Gomes and L. de Haan, *Approximation by penultimate extreme distributions*, Extremes **2**, 71-85 (1999).
- [25] L. de Haan and U. Stadtmüller, *Generalized regular variation of second order*, J. Austral. Math. Soc. **61**, 381 (1996).
- [26] G. Györgyi, N. R. Moloney, K. Ozogány, and Z. Rácz, *Finite-size scaling in extreme statistics*, Phys. Rev. Lett. **100**, 210601 (2008).
- [27] G. Györgyi, N. R. Moloney, K. Ozogány, Z. Rácz, and M. Droz, *Renormalization-group theory for finite-size scaling in extreme statistics*, Phys. Rev. E **81**, 041135 (2010).

- [28] S. M. Berman, *Limit Theorems for the Maximum Term in Stationary Sequences*, Ann. Math. Stat. **35**, 502 (1964).
- [29] J. F. Eichner, J. W. Kantelhardt, A. Bunde, and S. Havlin, *Extreme value statistics in records with long-term persistence*, Phys. Rev. E **73**, 016130 (2006).
- [30] S. N. Majumdar and A. Comtet, *Exact maximal height distribution of fluctuating interfaces*, Phys. Rev. Lett. **92**, 225501 (2004).
- [31] S. N. Majumdar and A. Comtet, *Airy distribution function: From the area under a Brownian excursion to the maximal height of fluctuating interfaces*, J. Stat. Phys. **119** 777-826, (2005).
- [32] G. Györfyi, N. R. Moloney, K. Ozogány, and Z. Rácz, *Maximal height statistics for $1/f^\alpha$ signals*, Phys. Rev. E **75**, 021123 (2007).
- [33] N. R. Moloney, K. Ozogány, and Z. Rácz, *Order statistics of $1/f^\alpha$ signals*, Phys. Rev. E **84**, 061101 (2011).
- [34] G. Schehr and S. N. Majumdar, *Universal Asymptotic Statistics of Maximal Relative Height in One-dimensional Solid-on-solid Models*, Phys. Rev. E **73**, 056103 (2006).
- [35] S. Sabhapandit and S. N. Majumdar, *Density of near-extreme events*, Phys. Rev. Lett. **98**, 140201 (2007).
- [36] T. W. Burkhardt, G. Györfyi, N. R. Moloney, and Z. Rácz, *Extreme statistics for time series: Distribution of the maximum relative to the initial value*, Phys. Rev. E **76**, 041119 (2007).
- [37] S. Redner, *A Guide to First-Passage Processes* (Cambridge University Press, Cambridge, 2001).
- [38] M. R. Leadbetter, G. Lindgren, and H. Rootzen, *Extremes and Related Properties of Random Sequences and Processes* (Springer-Verlag, New York, 1982).
- [39] S. Redner and M. R. Petersen, *On the Role of Global Warming on the Statistics of Record-Breaking Temperatures*, Phys. Rev. E **74**, 061114 (2006).
- [40] G. Wergen and J. Krug, *Record-breaking temperatures reveal a warming climate*, EPL **92**, 30008 (2010).
- [41] G. Wergen, M. Bogner, and J. Krug, *Record statistics for biased random walks, with an application to financial data*, Phys. Rev. E **83**, 051109 (2011).
- [42] J. Pickands, *Statistical Inference Using Extreme Order Statistics*, Annals of Statistics **3**, 119 (1975).

- [43] E. Brunet and B. Derrida, *Statistics at the tip of a branching random walk and the delay of traveling waves*, EPL **87**, 60010 (2009).
- [44] Y. V. Fyodorov, P. Le Doussal, and A. Rosso, *Statistical Mechanics of Logarithmic REM: Duality, Freezing and Extreme Value Statistics of $1/f$ Noises generated by Gaussian Free Fields*, J. Stat.Mech.: Theory and Exp., P10005 (2009).
- [45] T. Antal, M. Droz, G. Györgyi, and Z. Rácz, *Roughness distribution for $1/f^\alpha$ signals*, Phys. Rev. E **65**, 046140 (2002).
- [46] Haroon M. Barakat and Mouhamed A. El-Shandidy, *Order statistics with random sample size*, METRON - International Journal of Statistics, vol. LXII, n. 2, 233-246 (2004).
- [47] K. Ozogány, *Extrém érték statisztikák és fizikai alkalmazásai*, Diplomamunka (2006).
- [48] T. Antal, F. S. Labini, N. L. Vasilyev and Y. V. Baryshev, *Galaxy distribution and extreme-value statistics*, EPL, **88**, 59001 (2009).
- [49] D. G. York et al., *The Sloan Digital Sky Survey: Technical Summary*, AJ **120**, 1579 (2000).
- [50] C. Stoughton, R. H. Lupton, M. Bernardi, et al., *Sloan Digital Sky Survey: Early Data Release*, AJ, **123**, 485 (2002).
- [51] H. Aihara et al., *The Eighth Data Release of the Sloan Digital Sky Survey: First Data from SDSS-III*, ApJS **193** 29 (2011).
- [52] <http://casjobs.sdss.org>
- [53] M. Taghizadeh-Popp, *CfunBASE: A Cosmological Functions Library for Astronomical Databases*, PASP **122**, 976 (2010).
- [54] M. A. Strauss et al., *Spectroscopic Target Selection in the Sloan Digital Sky Survey: The Main Galaxy Sample*, AJ **124**, 1810 (2002).
- [55] D. J. Eisenstein et al., *Spectroscopic Target Selection for the Sloan Digital Sky Survey: The Luminous Red Galaxy Sample*, AJ **122**, 2267 (2001).
- [56] I. K. Baldry et al., *Quantifying the bimodal color-magnitude distribution of galaxies*, ApJ **600**, 681 (2004).
- [57] I. Csabai, A. J. Connolly, A. S. Szalay and T. Budavari, *Reconstructing Galaxy Spectral Energy Distributions from Broadband Photometry*, AJ **119**, 69 (2000).

- [58] K. M. Górski et al., *HEALPix: A Framework for High-Resolution Discretization and Fast Analysis of Data Distributed on the Sphere* ApJ **622**, 759 (2005).
- [59] <http://healpix.jpl.nasa.gov>
- [60] M. R. Blanton et al., *The luminosity function of galaxies in SDSS commissioning data*, The Astronomical Journal, **121** 2358-2380 (2001)
- [61] M. Taghizadeh-Popp, K. Ozogány, Z. Rácz, E. Regós, and A. Szalay, *Distribution of Maximal Luminosity of Galaxies in the Sloan Digital Sky Survey*, submitted to ApJ, arXiv:1204.0151 (2012).
- [62] W. H. Press and P. Schechter, *Formation of Galaxies and Clusters of Galaxies by Self-Similar Gravitational Condensation*, AJ **187**, 425 (1974).
- [63] T. Budavari et al., *Angular Clustering with Photometric Redshifts in the Sloan Digital Sky Survey: Bimodality in the Clustering Properties of Galaxies*, ApJ **595**, 59 (2003).
- [64] A. Dressler, *A comprehensive study of 12 very rich clusters of galaxies*, Ph.D. Thesis (1976).
- [65] S. D. Tremaine and D. O. Richstone, *A test of a statistical model for the luminosities of bright cluster galaxies*, ApJ **212**, 311 (1977).
- [66] Y.-T. Lin, J. P. Ostriker, and C. J. Miller, *A New Test of the Statistical Nature of the Brightest Cluster Galaxies*, AJ, **715**, 1486 (2010).
- [67] J. P. Ostriker and M. A. Hausman, *Cannibalism among the Galaxies: Dynamically Produced Evolution of Cluster Luminosity Function*, AJ, **217**, 125 (1977).
- [68] Y.-S. Loh, and M. A. Strauss, *The Bright End of the Luminosity Function of Red Sequence Galaxies*, Mon. Not. R. Astron. Soc., **366**, 373 (2006).
- [69] L. Dobos, and I. Csabai, *Order statistics of the early-type galaxy luminosity function*, Mon. Not. R. Astron. Soc., **414**, 1862 (2011).
- [70] A. Paranjape, and R. K. Sheth, *The luminosities of the brightest cluster galaxies and brightest satellites in SDSS groups*, arXiv:1107.3652v1
- [71] E. L. Scott, *The brightest galaxy in a cluster as a distance indicator*, AJ, **62**, 248 (1957).
- [72] S. P. Bhavsar, and J. D. Barrow, *First Ranked Galaxies in Groups and Clusters*, Mon. Not. R. Astron. Soc., **213**, 857 (1985).

ÖSSZEFOGLALÓ

Dolgozatunkban az extrém érték statisztikát véges és véletlen mintaméret mellett vizsgáltuk, foglalkoztunk korrelált és iid változók sorrend statisztikájával és galaxisok fényesség eloszlását elemeztük az extrém statisztika eszköztárával.

Bemutattuk az extrém határeloszlások és az első rendű véges méret korrekció meghatározásának reinterpretációját egy renormálási csoport bevezetésével. A Fisher-Tippett-Gumbel osztályban megadtuk a korrekció függvényének egzakt alakját az alkalmazások szempontjából lényeges sztenderdizációban.

Korrelált változók sorrend statisztikáját $1/f^\alpha$ zajokban vizsgáltuk, és a k -adik átlagos gap skálázása szerint három tartományt találtunk a korreláció függvényében. Megmutattuk, hogy iid változókra az első és második legnagyobb érték közötti gap határértéke függ az univerzalizációs osztálytól, és az asztrofizikai alkalmazás szempontjából fontos Fisher-Tippett-Gumbel osztályban meghatároztuk a gap véges méret korrekcióját.

Véletlen mintaméret mellett, ahol a végtelen mintaméret határesetét a mintaméret egy tipikus ($N_0 \rightarrow \infty$) értékével értelmezzük, megmutattuk, hogy az őseloszlás aszimptotikája szerinti szokásos univerzális osztályok megjelennek. Az extrém eloszlás három határeloszlás család egyikéhez fog tartani, s az adott családon belül további két esetet különböztetünk meg a mintaméret eloszlás skálázási tulajdonságai szerint. Az extrém határeloszlás az N_0 -al skálázó esetben függ a mintaméret eloszlás függvényalakjától. Az N_0 -al való skálázás során összeszűkülő mintaméret eloszlásnak nincs hatása az extrém határeloszlásra.

Egy adott térszögben megfigyelt galaxisok fényesség-eloszlását vizsgálva láttuk, hogy a fényességek extrém statisztikája az iid elmélet keretein belül értelmezhető. Jelentős véges méret korrekciót tapasztaltunk azonban mind a fényességek extrém eloszlásában, mind az első és második legnagyobb közötti távolság értékében. Mivel egy adott térszög galaxis-szám eloszlása keskeny volt, és N_0 -al összeszűkülött, ezért a fényességek extrém határeloszlása, a Schechter-őseloszlásból következően, a Fisher-Tippett-Gumbel-eloszlás volt, amely az empirikus eloszlásokat valóban jól leírta. Ennek a konklúzióknak a levonásához azonban a véges és véletlen mintaméretből adódó korrekciót is figyelembe kellett venni.

SUMMARY

Extreme value statistics is studied with emphasis on the effects of finite and random sample size, as well as on order statistics of correlated and iid variables. As an application we analyze the luminosity distribution of galaxies using the tools of extreme statistics.

We reinterpret the calculation of the limit distributions of extreme values in terms of renormalization group formalism. This approach allows a natural calculation of the leading order finite size correction to the limit distributions. In the case of the Fisher-Tippett-Gumbel (FTG) universality class we obtain the exact form of the correction function in the common standardization used in applications.

Order statistics of correlated variables is examined for $1/f^\alpha$ signals, and we find three scaling regimes for the average k th gap depending on the exponent α . We show that for iid variables the limiting value of the average gap between the first and second ranked values depends on the universality class, and for the FTG case we determine the finite size correction of the gap, which is an important quantity in astrophysics.

For the case of random sample size, where the limit of infinite sample size is defined through a typical value (N_0) of the sample size going to infinity, we show that the known universality classes emerge, according to the asymptotics of the parent. The extreme distributions can be classified into three limit classes, and inside a given class we can further distinguish two families according to the scaling properties of the sample size distribution. In the case when the sample size distribution scales with N_0 the limit distribution of the extremes depends on the functional shape of the sample size distribution. The sample size distribution does not influence the extreme limit distribution, if it shrinks sufficiently fast under the scaling.

Analyzing the empirical luminosity distribution of galaxies seen in a given solid angle of the sky we find that the extreme statistics of luminosities can be understood within the iid theory. On the other hand we observe large finite size corrections in the distribution of luminosity extremes, as well as in the gap between the first and second ranked luminosities. Since the galaxy count distribution in a solid angle is narrow and shrinks with N_0 , the limit distribution of extreme luminosities turns out to be the FTG distribution. In this picture we find excellent agreement with the empirical distributions but, in order to draw this conclusion we have to take into account the corrections arising from the finite and random sample size.